

OM of: US-09-367-496-8 to: GenEmbl.* out_format : pfs

Date: Jul 30, 2001 12:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=framet p2n.model -DEV=xlp
-O/cn2.1/USPTO.spool/US09367496/runat_30072001_081220_27373/app_query.fasta_1.637
-DB=GenEmbl -QFMT=fastap -SUPERIX=std.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QCAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-YGAPOP=6.000 -YCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09367496 -CGN1_1_5666 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPYX
-WAIT -THREADS=1

Search information block:

Query: US-09-367-496-8

Query length: 572

Database: GenEmbl.*

Database sequences: 1344157

Database length: -856060004

Search time (sec): 1878.380000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_p1:AB006713	+ 2872.00	3431.33	6.9e-183	2699	AB006713 Homo sapiens mRNA for
gb_p1:AB006716	+ 2882.00	3331.08	2.6e-177	1690	AB006716 Sequence 7 from Patent
gb_p1:AB006717	+ 2882.00	3331.08	2.6e-177	1690	AB006717 Sequence 7 from Patent
gb_p1:AB006718	+ 2882.00	3331.08	2.6e-177	1690	AB006718 Sequence 7 from Patent
gb_p1:AB006719	+ 2882.00	3331.08	2.6e-177	1690	AB006719 Sequence 7 from Patent
gb_p1:AB006720	+ 2882.00	3331.08	2.6e-177	1690	AB006720 Sequence 7 from Patent
gb_p1:AB006721	+ 2882.00	3331.08	2.6e-177	1690	AB006721 Sequence 7 from Patent
gb_p1:AB006722	+ 2882.00	3331.08	2.6e-177	1690	AB006722 Sequence 7 from Patent
gb_p1:AB006723	+ 2882.00	3331.08	2.6e-177	1690	AB006723 Sequence 7 from Patent
gb_p1:AB006724	+ 2882.00	3331.08	2.6e-177	1690	AB006724 Sequence 7 from Patent
gb_p1:AB006725	+ 2882.00	3331.08	2.6e-177	1690	AB006725 Sequence 7 from Patent
gb_p1:AB006726	+ 2882.00	3331.08	2.6e-177	1690	AB006726 Sequence 7 from Patent
gb_p1:AB006727	+ 2882.00	3331.08	2.6e-177	1690	AB006727 Sequence 7 from Patent
gb_p1:AB006728	+ 2882.00	3331.08	2.6e-177	1690	AB006728 Sequence 7 from Patent
gb_p1:AB006729	+ 2882.00	3331.08	2.6e-177	1690	AB006729 Sequence 7 from Patent
gb_p1:AB006730	+ 2882.00	3331.08	2.6e-177	1690	AB006730 Sequence 7 from Patent
gb_p1:AB006731	+ 2882.00	3331.08	2.6e-177	1690	AB006731 Sequence 7 from Patent
gb_p1:AB006732	+ 2882.00	3331.08	2.6e-177	1690	AB006732 Sequence 7 from Patent
gb_p1:AB006733	+ 2882.00	3331.08	2.6e-177	1690	AB006733 Sequence 7 from Patent
gb_p1:AB006734	+ 2882.00	3331.08	2.6e-177	1690	AB006734 Sequence 7 from Patent
gb_p1:AB006735	+ 2882.00	3331.08	2.6e-177	1690	AB006735 Sequence 7 from Patent
gb_p1:AB006736	+ 2882.00	3331.08	2.6e-177	1690	AB006736 Sequence 7 from Patent
gb_p1:AB006737	+ 2882.00	3331.08	2.6e-177	1690	AB006737 Sequence 7 from Patent
gb_p1:AB006738	+ 2882.00	3331.08	2.6e-177	1690	AB006738 Sequence 7 from Patent
gb_p1:AB006739	+ 2882.00	3331.08	2.6e-177	1690	AB006739 Sequence 7 from Patent
gb_p1:AB006740	+ 2882.00	3331.08	2.6e-177	1690	AB006740 Sequence 7 from Patent
gb_p1:AB006741	+ 2882.00	3331.08	2.6e-177	1690	AB006741 Sequence 7 from Patent
gb_p1:AB006742	+ 2882.00	3331.08	2.6e-177	1690	AB006742 Sequence 7 from Patent
gb_p1:AB006743	+ 2882.00	3331.08	2.6e-177	1690	AB006743 Sequence 7 from Patent
gb_p1:AB006744	+ 2882.00	3331.08	2.6e-177	1690	AB006744 Sequence 7 from Patent
gb_p1:AB006745	+ 2882.00	3331.08	2.6e-177	1690	AB006745 Sequence 7 from Patent
gb_p1:AB006746	+ 2882.00	3331.08	2.6e-177	1690	AB006746 Sequence 7 from Patent
gb_p1:AB006747	+ 2882.00	3331.08	2.6e-177	1690	AB006747 Sequence 7 from Patent
gb_p1:AB006748	+ 2882.00	3331.08	2.6e-177	1690	AB006748 Sequence 7 from Patent
gb_p1:AB006749	+ 2882.00	3331.08	2.6e-177	1690	AB006749 Sequence 7 from Patent
gb_p1:AB006750	+ 2882.00	3331.08	2.6e-177	1690	AB006750 Sequence 7 from Patent
gb_p1:AB006751	+ 2882.00	3331.08	2.6e-177	1690	AB006751 Sequence 7 from Patent
gb_p1:AB006752	+ 2882.00	3331.08	2.6e-177	1690	AB006752 Sequence 7 from Patent
gb_p1:AB006753	+ 2882.00	3331.08	2.6e-177	1690	AB006753 Sequence 7 from Patent
gb_p1:AB006754	+ 2882.00	3331.08	2.6e-177	1690	AB006754 Sequence 7 from Patent
gb_p1:AB006755	+ 2882.00	3331.08	2.6e-177	1690	AB006755 Sequence 7 from Patent
gb_p1:AB006756	+ 2882.00	3331.08	2.6e-177	1690	AB006756 Sequence 7 from Patent
gb_p1:AB006757	+ 2882.00	3331.08	2.6e-177	1690	AB006757 Sequence 7 from Patent
gb_p1:AB006758	+ 2882.00	3331.08	2.6e-177	1690	AB006758 Sequence 7 from Patent
gb_p1:AB006759	+ 2882.00	3331.08	2.6e-177	1690	AB006759 Sequence 7 from Patent
gb_p1:AB006760	+ 2882.00	3331.08	2.6e-177	1690	AB006760 Sequence 7 from Patent
gb_p1:AB006761	+ 2882.00	3331.08	2.6e-177	1690	AB006761 Sequence 7 from Patent
gb_p1:AB006762	+ 2882.00	3331.08	2.6e-177	1690	AB006762 Sequence 7 from Patent
gb_p1:AB006763	+ 2882.00	3331.08	2.6e-177	1690	AB006763 Sequence 7 from Patent
gb_p1:AB006764	+ 2882.00	3331.08	2.6e-177	1690	AB006764 Sequence 7 from Patent
gb_p1:AB006765	+ 2882.00	3331.08	2.6e-177	1690	AB006765 Sequence 7 from Patent
gb_p1:AB006766	+ 2882.00	3331.08	2.6e-177	1690	AB006766 Sequence 7 from Patent
gb_p1:AB006767	+ 2882.00	3331.08	2.6e-177	1690	AB006767 Sequence 7 from Patent
gb_p1:AB006768	+ 2882.00	3331.08	2.6e-177	1690	AB006768 Sequence 7 from Patent
gb_p1:AB006769	+ 2882.00	3331.08	2.6e-177	1690	AB006769 Sequence 7 from Patent
gb_p1:AB006770	+ 2882.00	3331.08	2.6e-177	1690	AB006770 Sequence 7 from Patent
gb_p1:AB006771	+ 2882.00	3331.08	2.6e-177	1690	AB006771 Sequence 7 from Patent
gb_p1:AB006772	+ 2882.00	3331.08	2.6e-177	1690	AB006772 Sequence 7 from Patent
gb_p1:AB006773	+ 2882.00	3331.08	2.6e-177	1690	AB006773 Sequence 7 from Patent
gb_p1:AB006774	+ 2882.00	3331.08	2.6e-177	1690	AB006774 Sequence 7 from Patent
gb_p1:AB006775	+ 2882.00	3331.08	2.6e-177	1690	AB006775 Sequence 7 from Patent
gb_p1:AB006776	+ 2882.00	3331.08	2.6e-177	1690	AB006776 Sequence 7 from Patent
gb_p1:AB006777	+ 2882.00	3331.08	2.6e-177	1690	AB006777 Sequence 7 from Patent
gb_p1:AB006778	+ 2882.00	3331.08	2.6e-177	1690	AB006778 Sequence 7 from Patent
gb_p1:AB006779	+ 2882.00	3331.08	2.6e-177	1690	AB006779 Sequence 7 from Patent
gb_p1:AB006780	+ 2882.00	3331.08	2.6e-177	1690	AB006780 Sequence 7 from Patent
gb_p1:AB006781	+ 2882.00	3331.08	2.6e-177	1690	AB006781 Sequence 7 from Patent
gb_p1:AB006782	+ 2882.00	3331.08	2.6e-177	1690	AB006782 Sequence 7 from Patent
gb_p1:AB006783	+ 2882.00	3331.08	2.6e-177	1690	AB006783 Sequence 7 from Patent
gb_p1:AB006784	+ 2882.00	3331.08	2.6e-177	1690	AB006784 Sequence 7 from Patent
gb_p1:AB006785	+ 2882.00	3331.08	2.6e-177	1690	AB006785 Sequence 7 from Patent
gb_p1:AB006786	+ 2882.00	3331.08	2.6e-177	1690	AB006786 Sequence 7 from Patent
gb_p1:AB006787	+ 2882.00	3331.08	2.6e-177	1690	AB006787 Sequence 7 from Patent
gb_p1:AB006788	+ 2882.00	3331.08	2.6e-177	1690	AB006788 Sequence 7 from Patent
gb_p1:AB006789	+ 2882.00	3331.08	2.6e-177	1690	AB006789 Sequence 7 from Patent
gb_p1:AB006790	+ 2882.00	3331.08	2.6e-177	1690	AB006790 Sequence 7 from Patent
gb_p1:AB006791	+ 2882.00	3331.08	2.6e-177	1690	AB006791 Sequence 7 from Patent
gb_p1:AB006792	+ 2882.00	3331.08	2.6e-177	1690	AB006792 Sequence 7 from Patent
gb_p1:AB006793	+ 2882.00	3331.08	2.6e-177	1690	AB006793 Sequence 7 from Patent
gb_p1:AB006794	+ 2882.00	3331.08	2.6e-177	1690	AB006794 Sequence 7 from Patent
gb_p1:AB006795	+ 2882.00	3331.08	2.6e-177	1690	AB006795 Sequence 7 from Patent
gb_p1:AB006796	+ 2882.00	3331.08	2.6e-177	1690	AB006796 Sequence 7 from Patent
gb_p1:AB006797	+ 2882.00	3331.08	2.6e-177	1690	AB006797 Sequence 7 from Patent
gb_p1:AB006798	+ 2882.00	3331.08	2.6e-177	1690	AB006798 Sequence 7 from Patent
gb_p1:AB006799	+ 2882.00	3331.08	2.6e-177	1690	AB006799 Sequence 7 from Patent
gb_p1:AB006800	+ 2882.00	3331.08	2.6e-177	1690	AB006800 Sequence 7 from Patent
gb_p1:AB006801	+ 2882.00	3331.08	2.6e-177	1690	AB006801 Sequence 7 from Patent
gb_p1:AB006802	+ 2882.00	3331.08	2.6e-177	1690	AB006802 Sequence 7 from Patent
gb_p1:AB006803	+ 2882.00	3331.08	2.6e-177	1690	AB006803 Sequence 7 from Patent
gb_p1:AB006804	+ 2882.00	3331.08	2.6e-177	1690	AB006804 Sequence 7 from Patent
gb_p1:AB006805	+ 2882.00	3331.08	2.6e-177	1690	AB006805 Sequence 7 from Patent
gb_p1:AB006806	+ 2882.00	3331.08	2.6e-177	1690	AB006806 Sequence 7 from Patent
gb_p1:AB006807	+ 2882.00	3331.08	2.6e-177	1690	AB006807 Sequence 7 from Patent
gb_p1:AB006808	+ 2882.00	3331.08	2.6e-177	1690	AB006808 Sequence 7 from Patent
gb_p1:AB006809	+ 2882.00	3331.08	2.6e-177	1690	AB006809 Sequence 7 from Patent
gb_p1:AB006810	+ 2882.00	3331.08	2.6e-177	1690	AB006810 Sequence 7 from Patent
gb_p1:AB006811	+ 2882.00	3331.08	2.6e-177	1690	AB006811 Sequence 7 from Patent
gb_p1:AB006812	+ 2882.00	3331.08	2.6e-177	1690	AB006812 Sequence 7 from Patent
gb_p1:AB006813	+ 2882.00	3331.08	2.6e-177	1690	AB006813 Sequence 7 from Patent
gb_p1:AB006814	+ 2882.00	3331.08	2.6e-177	1690	AB006814 Sequence 7 from Patent
gb_p1:AB006815	+ 2882.00	3331.08	2.6e-177	1690	AB006815 Sequence 7 from Patent
gb_p1:AB006816	+ 2882.00	3331.08	2.6e-177	1690	AB006816 Sequence 7 from Patent
gb_p1:AB006817	+ 2882.00	3331.08	2.6e-177	1690	AB006817 Sequence 7 from Patent
gb_p1:AB006818	+ 2882.00	3331.08	2.6e-177	1690	AB006818 Sequence 7 from Patent
gb_p1:AB006819	+ 2882.00	3331.08	2.6e-177	1690	AB006819 Sequence 7 from Patent
gb_p1:AB006820	+ 2882.00	3331.08	2.6e-177	1690	AB006820 Sequence 7 from Patent
gb_p1:AB006821	+ 2882.00	3331.08	2.6e-177	1690	AB006821 Sequence 7 from Patent
gb_p1:AB006822	+ 2882.00	3331.08	2.6e-177	1690	AB006822 Sequence 7 from Patent
gb_p1:AB006823	+ 2882.00	3331.08	2.6e-177	1690	AB006823 Sequence 7 from Patent
gb_p1:AB006824	+ 2882.00	3331.08	2.6e-177	1690	AB006824 Sequence 7 from Patent
gb_p1:AB006825	+ 2882.00	3331.08	2.6e-177	1690	AB006825 Sequence 7 from Patent
gb_p1:AB006826	+ 2882.00	3331.08	2.6e-177	1690	AB006826 Sequence 7 from Patent
gb_p1:AB006827	+ 2882.00	3331.08	2.6e-177	1690	AB006827 Sequence 7 from Patent
gb_p1:AB006828	+ 2882.00	3331.08	2.6e-177	1690	AB006828 Sequence 7 from Patent
gb_p1:AB006829	+ 2882.00	3331.08	2.6e-177	1690	AB006829 Sequence 7 from Patent
gb_p1:AB006830	+ 2882.00	3331.08	2.6e-177	1690	AB006830 Sequence 7 from Patent
gb_p1:AB006831	+ 2882.00	3331.08	2.6e-177	1690	AB006831 Sequence 7 from Patent
gb_p1:AB006832	+ 2882.00	3331.08	2.6e-177	1690	AB006832 Sequence 7 from Patent
gb_p1:AB006833	+ 2882.00	3331.08	2.6e-177	1690	AB006833 Sequence 7 from Patent
gb_p1:AB006834	+ 2882.00	3331.08	2.		

Align seg 1/1 to: AB006713 from: 1 to: 2699

```

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
149 ATGTCCTTCCAGGCAAGAAAGCATCCCGGGATCAGAGTACCGCCT 198
17 uLeuIleArgGlyArgIleValAlaAspGlnSerPheTyrAlaA 34
199 TCTGATCAGAGTGGAGATCGTGAATAGCAGCAGTCTTTACGCTG 248
34 spValHisValGluAspGlyLeuIleLysGlnIleGlyAlaSerLeu 50
249 ATGTGCACGTGGAGATGGCTGTGATAAAACAATCGGAGAAACCTCATC 298
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
299 GTCCCTGGGGCATCAAGACCATTCAGCCCGCCAGCCTGTATGGCTTCC 348
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
349 TGGTGGGCTTGTACCTGCACAGAGCTGCAGATGCCTGTCTGGGCATGA 398
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyCly 100
399 CACGGCTGAGGACTTCTGTGAGGGCACCAGGAGCGCTAGCAGGAGGA 448
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
449 ACCACCATGATCTTGGACCAAGCTTCTCCCGACACGGGTGTGAGCCTGCT 498
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
499 GCGGGCTCAGCGCTGGCGGGAGCGGGGAGCAGCGCGCTGTCTGGCG 548
134 spTyrSerLeuHisValAspIleThrArgThrHisGluSerIleLysGlu 150
549 ACTACTCCCTGCACGTGGACATCACCAGATGGCATGAGAGCATCAAGGAG 598
151 GluLeuGluAlaLeuValLysGlyLysValValAsnSerPheLeuValPh 167
599 GAGCTGGAGGCCCTGTGTCAGGAGAGAGGGTGTGAACCTCTCTCTGCTCT 648
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
649 CATGGCATACAGGACCGGTGCCAGTGCAGCAGCAGCAGATGTACGAGA 698
184 lePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
699 TCTTCAGCATCATCCGGACCTGGGGCCCTTGGCCAGCTGCACGCTGAG 748
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
749 AACGGGACATCGTGAGGAGGAGCAGAGAGCGGTGCTGGAGCTGGCAT 798
217 eThrGlyProGlyHisValSerHisProGluGluValGluAlaG 234
799 CACTGGCCCGAGGCGCAGTCTCAGCCACCCCGAGGAGGTGGAGGCTG 848
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
849 AGGCGGTGTACCGAGCTGTTCACCATCGCAGCAGCAGCAACTGCCCGCTG 898
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
899 TACGTACCAAGTGTATGAGCAGGGGGCGGCGGACCATCGCTCGCTCAGGC 948
267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
949 CAAGCGCAGAGGGTGGTGGTGGTGGGAGCCCATCACCAGCCAGCCTGG 998
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300

```

```

999 GCACCGACGGTTCACTACACTGAGGAGAAACTGGGCCCAAGGCCGACGCC 1048
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
1049 TTCGTCACATCACCCCTGTCAACCCAGACCCACCGCAGACACCT 1098
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
1099 CACCTGCTTGTGTCCAGCGGGACCTCCAGGTGACAGCAGCGCCCACT 1148
334 ystThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
1149 GCACCTTCACCACTGCCAGAGGCTGTGGCAGGACCACTTCGCGCTG 1198
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG1 367
1199 ATCCCGAGGCGCACACCGCATTTGAGGAGCGCATGTCGATGGTCTGGGA 1248
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
1249 GAAATGTGTGGCCTCTGGGAAGATGGAGCAGAGATGAGTTCGTCGCGTGA 1298
384 hrSerThrAsnAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
1299 CCAGTACAAATGCTGCCAAATCTTCAATTTTACCAAGGAAGGGCGGA 1348
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1349 GTGGCTGTGGCTCTGACGCTGACCTGTCATATGGAACCCCAAGGCCAC 1398
417 rLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1399 CAGATCATCTCTGCCAAGACCCCAATCTGAAGTGGAGTACAACTCT 1448
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1449 TCGAGGAGTGGAGTCCGGGGAGCGCTCCCGTGGTGCATTAAGTCAGGC 1498
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1499 CGAGTGGCGCTGGAGGACGGGAAGATGTTTGTACCCCGGGCGCGCGC 1548
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1549 CTTCTGCTCCCTCGGAAACATTCCTCCGACCTTGTCTACAAGAGGATCAA 1598
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1599 CTCGCAACAGGCTGGCGGAGATCCACGGTGTGCCCGTGGGCTGTATGAC 1648
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1649 GGGCGCTCCAGAGTGTATGTCCTGCCAAGCCAGGAGTGGCGCTCC 1698
517 alaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
1699 GGCCGCGCGCTCTGCCAGGCAAGATCTCCGTGCTCTCTCTCTCTCTCT 1748
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
1749 TACATCAGTCGGGGTTCAGCCTTATCTGGGTCTCAGGCTGATGACCACATC 1798
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
1799 GCCCGACGCACAGCAGAGATCATGGCACCCCTGGCGGCGCGCTCCAA 1848
567 nIleThrSerLeuSer 572
1849 CATCACCTCTCTCTCC 1864
seq_name: gb_pat1:A87376
seq_documentation_block:

```

LOCUS A87376 1690 bp DNA PAT 22-JAN-2000
 DEFINITION Sequence 7 from Patent WO9837192.
 ACCESSION A87376
 VERSION A87376.1 GI:6736129
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 1690)
 AUTHORS Byk, T. and Belin, M.
 TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
 PARANEOPLASTIC NEUROLOGICAL SYNDROMES
 JOURNAL Patent: WO 9837192-A 7 27-AUG-1998;
 BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
 FEATURES
 Location/Qualifiers
 source 1..1690
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 360 a 505 c 539 g 286 t
 ORIGIN
 alignment_scores:
 Quality: 2882.00 Length: 553
 Ratio: 5.221 Gaps: 0
 Percent Similarity: 99.819 Percent Identity: 99.819
 alignment_block:
 US-09-367-496-8 x A87376 ..
 Align seg 1/1 to: A87376 from: 1 to: 1690
 1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 |||||
 32 ATGTCCTCCAGGCAAGAAAGATCCCGCGGATCAGAGTGACCGCCT 81
 |||||
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
 |||||
 82 TCTGATCAGAGGTGGGAGGATCGTAATGACGACCATCTCTTTAGCTG 131
 |||||
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 |||||
 132 ATGTCAGCGTGAAGATGGCTTCATAAAACAAATCGAGAAACCTCATC 181
 |||||
 51 ValProGlyLysIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 |||||
 182 GTCCCTGGGGGCATCAGACCATGACGCCCGCGCTGATGCTCCTTCC 231
 |||||
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 |||||
 232 TGGTGGCGTTGAGGTCACACAGAGGTGCAGATGCCGTCTCTGGCATGA 281
 |||||
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaLeuAlaGlyGly 100
 |||||
 282 CACCGCGTGACGACTTCTGTGAGGCGACCAAGGCGCGCTAGCAGAGGA 331
 |||||
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 |||||
 332 ACCACCATGATCTGGACCACTCTTCCCGACCGCGGTGTGAGCTGCT 381
 |||||
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 |||||
 382 GCGCGCTACGAGCAGTGGCGGAGCGCGGCGAGCGCGCGCTGCTGCG 431
 |||||
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 |||||
 432 ACTACTCCCTGCAGTGGACATCATCCCGATGGCATGAGACATCAAGGAG 481
 |||||
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 |||||
 482 GAGCTGGAGGCGCTGGGTCAAGGAGAGGGGTGTGAACCTCTCTCTGGTCTT 531
 |||||

167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 |||||
 532 CATGGCATACAAGGACCGGTGCCAGTGCAGCGACGACGACGATGTACGAGA 581
 |||||
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 |||||
 582 TCTTTCAGCATCATCGGACCTGGGGGCTTGGCCAGGTGCACGCTGAG 631
 |||||
 201 AsnGlyAspIleValGluGluGluLysArgLeuLeuGluLeuGlyIle 217
 |||||
 632 AACCGGACATCTGTGGAGGAGGACGAGCGGTGCTGGAGCTCGGCAT 681
 |||||
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 |||||
 682 CAC"GGCCCGGAGGCGGCGCTGCTCAGCCACCCCGAGGAGGTGAGGCTG 731
 |||||
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 |||||
 732 AGGGGTGTACCGAGCTGTCCACCATCGCAAGCAGGCAAACTGCCCGCTG 781
 |||||
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 |||||
 782 TACGTCCACCAAGGTGTAGAGCAAGGGCGCGCGCCATCGCTCAGGC 831
 |||||
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
 |||||
 832 CAAGCGCAGAGGGGTGCTGCTGTTGGGAGCCCATCACCGCAGCCTGG 881
 |||||
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaA 300
 |||||
 882 GCACCGCAGCGGTTCACACTACTGGAGCAAGAACTGGGCGCAAGCTGCAGCC 931
 |||||
 301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
 |||||
 932 TTCCTCATCATCACCCCTGTCAACCCAGACCCACCGCAGACCACT 981
 |||||
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 |||||
 982 CACCTGTCTGTCTCCAGCGGGACCTCCAGGTGACAGGCGAGCCCACT 1031
 |||||
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 |||||
 1032 GCACCTTCACCACTGCCAGAAAGGCTGTGGGCAAGGACAACTTCGCGCTG 1081
 |||||
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 |||||
 1082 ATCCCGGAGGCGACCAACCGCATTTGAGGAGCGCATGTCGATGCTGCGGA 1131
 |||||
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 |||||
 1132 GAATGTGTGGCTCTGGGAAGATGGACGAGATGAGTTCTGTCGCGGTGA 1181
 |||||
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 |||||
 1182 CCAGTACAAATCTGCCAAATCTCAATTTTACCCAAAGGAAGGGCGCA 1231
 |||||
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 |||||
 1232 GTGGCTGTGGCTCTGACGCTGACCTGGTGCATATGGAAACCCCAAGGCCAC 1281
 |||||
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 |||||
 1282 CAAGATCATCTCTGCCCAAGACCCCACTCTGAACGTGGAGTACACATCT 1331
 |||||
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 |||||
 1332 TCCAGGAGGTGGAGTCCCGGGAGCGCTGCGCTGCTCATAGTCAGGCGC 1381
 |||||
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyArg 467
 |||||
 1382 CGAGTGGCGCTGGAGGACGGGAAGATGTTTGTACCCCGGGGGCGGCGG 1431
 |||||
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 |||||

1432 CTTGTCCTCGGAACATCCCGGACTTTGCTACAAGAGGATCAAG 1481
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1482 CTCGCAACAGGCTGGCGAGATCCACGGTGTGCCCGTGGCTGTATGAC 1531
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1532 GGGCCGCTCCACGAGGTGATGGTGCCTGCCAAGCCAGGAGTGCGCTCC 1581
 517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
 1582 GGCCCGCGCTGCTGCCAGGCAAGATCTCCGTGCTCTGTGCGCAACC 1631
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
 1632 TACATCAGTCGGGTTCAGCTATCTGGGTCTCAGGCTGATGACCATC 1681
 551 AlaArgArg 553
 1682 GCGCGAGCG 1690

seq_name: gb_pr9:HSRNAULI4

seq_documentation_block: 1690 bp mRNA PRI 09-MAY-1997
 LOCUS HSRNAULI4
 DEFINITION H.sapiens mRNA for Ulip4 protein.
 ACCESSION Y10976
 VERSION Y10976.1 GI:2077995
 KEYWORDS phosphoprotein.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1690)
 By, T., Ozon, S. and Sobel, A.
 The Ulip4s: a family of proteins related to the axonal guidance
 associated unc-33 gene product
 Unpublished
 2 (bases 1 to 1690)
 Direct Submission
 Submitted (03-FEB-1997) T. Byk, INSERM U440, 17 rue du Fer a
 Moulin, F- 75005 Paris, FRANCE
 On May 10, 1997 this sequence version replaced gi:1914866.
 Related sequence X87817.
 Location/Qualifiers
 1..1690
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="retina"
 /clone="ATCC No. 469253"
 32..>1690
 /note="mistake in reverse transcription 197..199"
 /codon_start=1
 /product="cytosolic phosphoprotein"
 /protein_id="CAA71872.1"
 /db_xref="GI:1946203"
 /db_xref="SWISS-PROT:O14531"
 /translation="MSFOGKKSIPRTSDRLIRGRIVNDQSFYADVHVEDLIQK
 IGNLIVPGGIXITIDAGMLVLPQGVVHTRLQPMVLTGTPADDFCGTKAALAGGTT
 MILDHVPPTGVSLLAAYEQRERADSAACDYSLHVDITRWHEISKEELALVKEKG
 VNSFLVMAYKDRCCSDSQMYEFSIIRDLGLAQVHAENGDIIVEEOKRRLLELIGT
 GPBGHVLSPHEVEAEAVRYNTIAQANCLYTKVNSKGAADAIAQAKRRGVVVFV
 EPTASLSDGDSHYNSKNAKAAAFVTSPPNPDPTADHLCCLSSGLQVTVSAHC
 TTTAQAKGKNFALIPETNGIEERMSMWEKCVASGMDENEFVAVTSTNAKIF
 NFYPRKRVAVGSDADLVITWNPATKIIISAKTHNLNVEYNIFEGVRCGAPAVVISOG
 RVALEDKMFVTPGAGRFVPRKTFPDPVYKRIKARNLAEIHGVPRLYDGPVHEVMV
 PAKPGSGAPARASCPKISVPPVPRNLHQSGFSLSGQADDIARR"
 359 a 505 c 539 g 286 t 1 others

CDS

167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 482 GAGCTGGAGGCCCTGGTCAAGAGAGAGGGGTGTGAATCCTCTCTGCTCTT 531
 532 CATGCGATACAAGAGCCGGTGCAGTGCAGCCAGCACGCCAGATGTACGAGA 581
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 582 TCTTCAGCATCATCCGGGACCTGGGGGCGCTTGGCCAGGTGCACGCTGAG 631
 201 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuGluLeuGlyI 217
 632 AACGGGGACATCGTGGAGGAGGAGGAGGAGCGGTGTGCTGGAGCTCGGCAT 681
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 682 CACTGGCCCCGAGGGCCACGTGCTCAGCCACCCCGGAGGAGGTGGAGGCTG 731
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 732 AGGGCGGTACCGAGCTGTACCATCGCAAGAGGAGCAAACTGCCCGCTG 781
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267

alignment_scores:
 Quality: 2882.00 Length: 553
 Ratio: 5.221 Gaps: 0
 Percent Similarity: 99.819 Percent Identity: 99.819
 alignment_block:
 US-09-367-496-8 x HSRNAULI4 ..
 Align seg 1/1 to: HSRNAULI4 from: 1 to: 1690
 1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
 32 ATGTCCTTCCAGGCAAGAAAGCATCCCCGGATCACGAGTGACCGCT 81
 17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
 82 TCTGATCAGAGGTGGGAGGATCGTGAATGACGACCATCTTTTACGCTG 131
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 132 ATGTGACGTGGAGATGGCTTGATAAAACAAATCGGAGAAACCTCATC 181
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 182 GTCCCTGGGGGCATCAGACCATTCAGCGCCACCGCTGATGGTCTCTTC 231
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 232 TGGTGGCGTTGACGTCCACACAAGGCTGCAGATGCCTCTCTGGGCATGA 281
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 282 CACCGGCTGACGACTTCTGTACGGGACCAAGCAGCGCTAGCAGGAGGA 331
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
 332 ACCACCATGATCTTGACACACGCTTTCCTCCGACACGGGTGTGAGCTGCT 381
 117 uAlaAlaTyrGluThrArgGluArgAlaAspSerAlaAlaCysCysA 134
 382 GCGGGCTACGAGCAGTGGGGAGGGGGGAGCGGGGACAGCGGGCGCTGTCGG 431
 134 spTyrSerLeuHisValAspIleThrArgThrHisGluSerIleLysGlu 150
 432 ACTACTCCCTGCAGCTGGACATCACCCGATGCGATGAGAGCATCAAGGAG 481
 151 GluLeuGluAlaLeuValLysGlnLysGlyValAsnSerPheLeuValPh 167
 482 GAGCTGGAGGCCCTGGTCAAGAGAGAGGGGTGTGAATCCTCTCTGCTCTT 531
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 532 CATGCGATACAAGAGCCGGTGCAGTGCAGCCAGCACGCCAGATGTACGAGA 581
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 582 TCTTCAGCATCATCCGGGACCTGGGGGCGCTTGGCCAGGTGCACGCTGAG 631
 201 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuGluLeuGlyI 217
 632 AACGGGGACATCGTGGAGGAGGAGGAGGAGCGGTGTGCTGGAGCTCGGCAT 681
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 682 CACTGGCCCCGAGGGCCACGTGCTCAGCCACCCCGGAGGAGGTGGAGGCTG 731
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 732 AGGGCGGTACCGAGCTGTACCATCGCAAGAGGAGCAAACTGCCCGCTG 781
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267

Tue Jul 31 13:08:18 2001

```

782 TACGTACCAAGGTGATGACCAAGGGGGGGGGCCAGCCATCGCTCAGGC 831
267 alyArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
832 CAAGCGCAGAGGGTGGTGGTGGTGGGAGCCCATCCGCCAGCGCTGG 881
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
882 GCACCGACGGTTCACACTTACCTGGAGCAGAACTGGCCCAAGCTGCAGCC 931
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
|||||
932 TTCGTACATCACCCTGTCAACCCAGACCCACCCAGCGCAGACCACT 981
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
982 CACCTGCTGTGTCCAGCGGGAGCTCCAGGTGACAGCGAGCGCCACT 1031
334 ysThrPheThrThrAlaGlnLysAlaValAlGlyLysAspAsnPhaAlaLeu 350
|||||
1032 GCACCTTACCCTGCCAGAGAGCTGTGGCAAGGACCACTTCGCGCTG 1081
351 ileProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpL 367
|||||
1082 ATCCCGAGGCGCACCAACGCGATTCAGGAGCGCATGTCGATGCTGGGA 1131
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1132 GAAATGTGGGCTCTGGGAGAGTGCAGGAGTATGAGTTCGTCGCGGTGA 1181
384 hrSerThrAsnAlaAlaLysIlePheAsnPhetYrProArgLysGlyArg 400
|||||
1182 CCAGTACAAATGTCGCAAAATCTCAATTTTACCCAGGAGGCGGA 1231
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1232 GTGGCTGTGGGCTCTACGCTGACCTGGTCTATGATGACCAAGCGCAC 1281
417 rLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1282 CAAGATCATCTCGCAAGACCCACCAATCTGAACGTGGAGTACAACTCT 1331
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1332 TCGAGGAGTGGAGTGGCGGGAGCGCTGCGGTGGTCAATAGTCAGGCG 1381
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1382 CGAGTGGCGCTGGAGGAGCGGGAAGATGTTGTCAACCCCGGGGGCGCG 1431
467 qPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1432 CTCTGCTCCCTCGGAAACATTCGCGGAGTGTGTCTACAAGAGGATCAAG 1481
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1482 CTCGCAACAGGCTGGCGGAGATCCACGCTGCCCCGTGGCTGTATGAC 1531
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1532 GGGCCCGTCCAGGAGTGTGCTGCTGCCAAGCCAGGAGTGGCGCTCC 1581
517 alaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1582 GGCCCGCGGCTCCTGCCAGGCAAGATCTCCGCTGCTCTCTGTCGCCAAC 1631
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
|||||
1632 TACATCAGTCGGGTTCAGCCTATCTGGGTCTCAGGCTGATGACCATC 1681
551 AlaArgArg 553
|||||
1682 GCCCGACGC 1690

```

seq_name: gb_rol:AB006715

seq_documentation_block:
LOCUS AB006715 2546 bp mRNA ROD 23-AUG-1997
DEFINITION Mus musculus mRNA for dihydropyrimidinase related protein 4,
complete cds.

ACCESSION AB006715.1 GI:2342489
VERSION dihydropyrimidinase related protein 4.
KEYWORDS Mus musculus fetus whole embryo cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2546)
AUTHORS Nonaka,M.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1997) to the DDBJ/EMBL/GenBank databases. Masaru
Nonaka, Nagoya City University Medical School, Department of
Biochemistry; 1 Kawasumi, Mizuho-cho, Mizuho-ku, Nagoya, Aichi 467,
Japan (E-mail:mnonaka@med.nagoya-cu.ac.jp, Tel:052-853-8145,
Fax:052-842-3955)

REFERENCE 2 (sites)
AUTHORS Hamajima,N., Kato,Y., Kowaki,M., Wada,Y., Sasasaki,M. and Nonaka,M.
TITLE Novel members of dihydropyrimidinase related protein family
JOURNAL Unpublished (1997)
REFERENCE 3 (sites)
AUTHORS Hamajima,N., Matsuda,K., Sakata,S., Tamaki,N., Sasaki,M. and
Nonaka,M.
TITLE A novel gene family defined by human dihydropyrimidinase and three
related proteins with differential tissue distribution
JOURNAL Gene 180 (1-2), 157-163 (1996)
MEDLINE 97128821

FEATURES
source 1..2546
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="fetus"
/tissue_type="whole embryo"
52..1770
/codon_start=1
/product="dihydropyrimidinase related protein 4"
/protein_id="BAA21888.1"
/db_xref="GI:2342490"
/translation="MSFGKKSIPIRITSDRLIKGGKIVNDQSFHADLYVEDGLIKQ
IGENLYPGGKTIIDAHGLVLPGLVDVHTRQLQPNVGMTPADPCOGTKAALAGGTT
MILDRVFDAGVSLAAAYEQWRERADSAACDYLSDVDPKRWHESTKEELEALVRDKG
VNSFLVFMAYKDRCCQTDGQIYEIFSLRIDLGAQAQVHAENGDIIEEQRKLLQGLIT
GPEGHVLSHPVEAEAYRAVTAKOANCPYTVKVMKGAADVAQAKRRGVVFG
EPITASLGTDGSHYWSKNWAKAAAFVTPPINPDPTTADHLTSLSSGDLQVTSAGHC
TFTTAQKAVGKDNFTLIPGEGVNGIEERMSVYVEKVASGKMDENEFVAVTSTNAKIF
NFYPRKGRVAVGSDADLVINPRAIKVFSKSHNLNVEYNIIEGVECRGVPTVVISOG
RVVLEDGNLVTGAGREIFPKTFFDFYKRIKARNLAEIHGVRGLYDGPVHEVML
PAKPGSGTQARASCSGKISVPPVRNLHQSGLSGSQADDDHIARTAKIIMAPGGRS
NITSLS"

polyA_site 2546
BASE COUNT 558 a 719 c 738 g 531 t
ORIGIN
alignment_scores:
Quality: 2818.00 Length: 572
Ratio: 4.996 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 93.007

alignment_block:
US-09-367-496-x AB006715 ..
Align seg 1/1 to: AB006715 from: 1 to: 2546

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
|||||

52 ATGTCCTTCCAAAGCAAGAGATTCCTCCGGATAAACGAGCGCGCT 101
 17 uLeuileArgGlyGlyArgGlyValAlaAsnAspGlnSerPheThrAlaA 34
 102 TCTCATCAAGAGTGGGAAGATGTGACCATGACCATCTTTTCATGTG 151
 34 spValHisValGluAspGlyLeuileLysGlnIleGlyGluAsnLeuile 50
 152 ATCTGTATGTGAAGAGCGTCTGATTAAACAAATTTGGAGAAATCTCATC 201
 51 ValProGlyGlyLysHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 202 GTCCTTGGGGGATCAAAACCATTCATGCTATGCGCTGATGGTCTGCC 251
 67 GglyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
 252 TGGGGAGTTGACGTTACACCCGCTCAGATGCTGTGATGGCATGA 301
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 302 CCCAGCTGATGATTTCTGTCAGGCGACCAAGCGGCTCTAGCAGGGGG 351
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerIleuLe 117
 352 ACCACCATGATATTGGACCATGTGTTCTTCCAGCCTGGTGTGACCTTGC 401
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 402 GGCAGCTTATGAGCAGTGGGGAGCGAGCAGACAGCGCGCTGCTGTG 451
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 452 ACTACTCTTACATGTGACATTCCTCGCTGGCAGCAGACCAAGAA 501
 151 GluLeuGluAlaLeuValLysGluLysGlyValAlaAsnSerPheLeuValPh 167
 502 GAGCTGAGGCGCTTAGTCAGGCAAGAGTGTGAATCTCTTCTGCTCT 551
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 552 CATGGCATACAAGACAGAGTGGCAGTGTACTGACGCTGATATATGAA 601
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 602 TCTTCAGCTCTATCCGGGACCTGGAGTGTGGCCAGGTGACACAGAA 651
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLysIle 217
 652 AATGGGACATCTGGAGGAGNACAGAGCGCTGCTGGAGCAAGGCAT 701
 217 eThrGlyProGlyHisValLeuSerHisProGluGluValGluAlaG 234
 702 CACTGTCTGAGGGCCATGTGCTCAGCCACCCAGAGAGGTAGAGCGG 751
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 752 AGGCTGTGTACAGACAGTCACTATGTCAGAGCAGGCAACTGCCACTA 801
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 802 TACGTACCAAGGTGATGAGAAGGTGCAGCTGACATGGTGTGCCAAGC 851
 267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
 852 CAAGCGCAGGGGGTGGTCTTTTGGGAACCTATCATGCCAGCTGG 901
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 902 GCACGTGATGGCTCACACTACTGGAGCAGAGACTGGGCCAAGGTGACGC 951
 301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
 952 TTTGTCACTTCAACCCCTATCAACCCGAGCCCTACTACTGCAGACCACT 1001

317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1002 CACTCTCTGTCTGTCAGTGGGACCTCCAGGTGACAGGAGTGCACCT 1051
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1052 GCACCTTCTACTACTGCCAGAGCGCTTGGCAAGACAACTTCACACTG 1101
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1102 ATCCCCGAGGAGTCAACGGTATAGAGAGCGCATGTCTGTGCTGGGA 1151
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 1152 GAAATGTGTGGCTTACAGGGAATGACAGAGATGAGTTCGTGGCGTGA 1201
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1202 CCAGCAAAATGCTGCCAAATCTTCAATTTTACCCAGGAAGGGCGT 1251
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1252 GTGGCGTGGCTCTGATGCTGACCTGGTCACTGGACCCCGAGGCCAC 1301
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1302 GAAAGTCTTCTGTCAGAGAGCAATCACTGAATGTAGAGTACAACATCT 1351
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 1352 TTGAGGAGTGGAGTGGCGAGGAGTGGCCAGCGTGGTCAATAGTCAGGC 1401
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1402 AGAGTGGTGTGGAGGACGAAACCTGCTTGTACTCCAGGGGCTGGCGG 1451
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLys 484
 1452 CTTCATCCCGGAGACGTTCCCGGACTTTGTCTATAAGAGAGATAAGG 1501
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1502 CTCGACACAGGCTAGCAGAGATCCACGGTGTGCTCCAGGCGCTGACGAC 1551
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1552 GGGCTGTGCATGAAGTGTGTTACCTGCCAAGCCAGGAAGTGGCACACA 1601
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1602 GGCCGCTGCATCTGTTTCAGGCAAGATCTCAGTGCCACCCGTCGCAACC 1651
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
 1652 TGACCCAGTGGGGTTCAGCTATCTGGCTCTCAGGCTGACGATCATT 1701
 551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
 1702 GCCAGAGGTACGGCTCAGAGATCATGGCACCCCGGAGGAGCGCTCCAA 1751
 567 nIleThrSerLeuSer 572
 1752 CATCAGCTCTCTTTC 1767

seq_name: gb_pat1:A87374

seq_documentation_block:

LOCUS A87374 1920 bp DNA

DEFINITION Sequence 5 from Patent WO9837192.

ACCESSION A87374

VERSION A87374.1 GI:6736128

KEYWORDS

PAT

22-JAN-2000

Tue Jul 31 13:08:18 2001

us-09-367-496-8.std.rge

SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 1920)
 AUTHORS Byk,T. and Belin,M.
 TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
 PARANEOPLASTIC NEUROLOGICAL SYNDROMES
 JOURNAL Patent: WO 9837192-A 5 27-AUG-1998;
 BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
 FEATURES Location/Qualifiers
 source 1..1920
 /organism="Mus musculus"
 /db_xref="taxon:10090"

BASE COUNT 440 a 524 c 562 g 394 t
 ORIGIN

alignment_scores:
 Quality: 2803.00 Length: 572
 Ratio: 4.988 Gaps: 0
 Percent Similarity: 98.252 Percent Identity: 92.657

alignment_block:

US-09-367-496-8 x A87374 ..

Align seg 1/1 to: A87374 from: 1 to: 1920

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLeu 17
 102 ATGTCCTTCCCAAGGCAAGAGAGCATCCCCGGATACAGGACCGGCT 151
 17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
 152 TCTCATCAAGGTGGGAAGATTGTGAACGATGACCACTCTTTTCATGCTG 201
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyClnAsnLeuLe 50
 202 ATCTGTATGTGAAGAGCGTCTGATTAAACAATTTGGAGAAATCTCATVC 251
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 252 GTCCCTGGGGGCATCAAAACCATCGATGCTCATGGCTGATGGTGTGCC 301
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 302 TGGGGAGGTGACCTTCACACCCGGCTGCAGATGCCTGTGATGGGCATGA 351
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCAGCTGATGATTCTGTGCAGGGCACCACCGCGCTCTAGCAGGGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 402 ACCACCATGATATGGACCATGTGTTCTTCAGCGCTGGTGTGAGCTGCT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 452 GGCAGCCTATGACAGGTGGCGGGAGGAGCAGACAGCGCGGCTGCTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 502 ACTACTCTTACATGTGGACATCTCTCGCTGGCAGCAGACCAACAAAGAA 551
 151 GluLeuGluAlaLeuValLysGluGlyGlyValAsnSerPheLeuValPh 167
 552 GAGCTGGAGGCCCTAGTCAGGGACAAAGGTGTGAACCTCTCTCTGGTCTT 601
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 602 CATGGCATACAAAGGACAGTCCAGTGTACTGACGGGTGACAGATATGAAA 651
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200

652 TCTTCAGCTCATCTCCGGGACCTGGGAGCTGTGGCCAGGTGCACGCAGAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
 702 AATGGGACATCTGTGAGGAGGAACAAGACCGCTGTCTGGAGCAAGGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 752 CACTGGTCTCTGAGGGCCATGTGTCTCAGCCACCACCAAGAGGTAGAGCGG 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 802 AGGCTGTGTACAGCAGCTCACCATTGCCAAGCAGGCCAACTGCCCACTA 851
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 852 TAGCTCACCAAGGTGATGAGCAAGGGTGCAGCTGACATGGTTGCCCAAGC 901
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
 902 CAACCGCAGGGGGTGTCTCTTTGGGAACTTATCCTGCGCAGCCACTGG 951
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaAla 300
 952 GCACCTGATGGCTCACACTACTGTGAGCAAGAACTGGGCCAAGGCTGCAGCC 1001
 301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
 1002 TTTGTCACTTCCACCCCTCATCAACCCGACCCCTACTACTGCGACCCACT 1051
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1052 CACTCTCTGCTGTCCAGTGGGGACCTCCAGGTGACAGGCAGTGCCTACT 1101
 334 ystThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1102 GCACCTTCACTACTGCCCAAGAGGCTGTGGCAAGACAACTTCACACTG 1151
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1152 ATCCCCGAGAGTAGTCAACGGTATAGAACGGCGCATGTCTGTGTCTGGGA 1201
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 1202 GAAATGTGTGGCTTCAGGGAAATGGACGAGAAATGAGTTCTGTTGCCGTGA 1251
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1252 CCAGCACAAATGCTGCCAAATCTCAATTTTACCCAGGAAGGGCGT 1301
 401 ValAlaValIcYSerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1302 GTGCGCTGGCTCTGATCTGACCTGGTCTATCGAACCCAGGCGCCAC 1351
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1352 GAAGTCACTCTGTCCAGAGCATTAACCTGAATGTAGAGTACACACTCT 1401
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 1402 TTCAAGAGGTGGAGTCCCGAGGAGTGCCACGGTGGTGTATAGTCAAGGC 1451
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1452 AGAGTGTGTGGAGGACGAAACCTGCTGTCTCCTCCTCCAGGGGCTGGCCG 1501
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 1502 CTTCATTCCCCGGAAGACGTTCCCGGACCTTTGTCTATAGAGCATAAAGG 1551
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500

1552 CTCGACAGAGCTAGCAGAGATCCAGGTGTCCTCGAGGCTGTACGAC 1601
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1602 GGGCTGTGCATGAAGTATCTACCTGCCAAGCCAGGAGTGGCACACA 1651
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1652 GGCCCGGTGCATCTGTTTCAGCAAGATCTCAGTGCACCCCGTGGCAACC 1701
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
 1702 TGCACCACTCGGGTTCAGCTACTCTCAGGTCTGAGGTGAGATCACTT 1751
 551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
 1752 GCCACAGGTACGGCTCAGAAGATCATGGACCCCGGAGGACCTCAA 1801
 567 nileThrSerLeuSer 572
 1802 CATCACGTCTCTTCC 1817

seq_name: gb_rol:MMULIP4

seq_documentation_block: 1920 bp mRNA ROD 18-APR-1997
 LOCUS MMULIP4
 DEFINITION M.musculus mRNA for Ulip4 protein.
 ACCESSION Y09079
 VERSION Y09079.1 GI:1915916
 KEYWORDS phosphoprotein; Ulip4 gene.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1920)
 Byk,T.; Ozon,S. and Sobel,A.
 The Ulips: a family of proteins related to the axonal guidance
 associated unc-33 gene product
 Unpublished
 2 (bases 1 to 1920)
 Byk,T.
 Direct Submission
 Submitted (28-OCT-1996) T. Byk, INSERM U440, 17 rue du Fer a
 Moulin, F- 75005 Paris, FRANCE
 Related sequence: X87817.
 Location/Qualifiers
 1. 1920
 /organism="Mus musculus"
 /sub_species="Harlan Sprague Dawley"
 /strain="ICR outbred strain"
 /db_xref="taxon:10090"
 /dev_stage="newborn"
 /tissue_type="brain"
 /clone_lib="lambda Zap II, Stratagene, # 936309"
 102..1820
 /gene="Ulip4"
 102..1820
 /genes="Ulip4"
 /codon_start=1
 /evidence="experimental"
 /product="Ulip4 protein"
 /protein_id="CAA70299.1"
 /db_xref="GI:1945273"
 /db_xref="SWISS-PROT:O35098"
 /translation="MSFQKKSIPIRTSDRLIKGKIVNDQSFHADLYVDDGLIKQ
 IGENIVPGIKTIDAGLMVLPGGVDVHTLQPMVMTPADDFCQGTAKAALAGTT
 MILDHFDPAGVSLIAAEQWRDGSACCDYSLHVDIPRWHESTKEELEALVRDKG
 VNSLFLVWAKDRCCDQGYEIFESLIRDLGAQVHAENGDIIVEEQKRLLEGIT
 GPESHLVSHPEVEAEAYRAVTIAQANCPLYTVKMSKGAADMAVQAKRRGVVVF
 EPTASLTDGDSHWKSKWAKAAAFVTSPPINPDPTADHLTSLSSGLDQVTSAGHC
 TTTTAQAKGNDFTLIPVWNGIEERMSVWVKCVASGKMDNEFEVAVTSNAKIF
 NFYPRKRVAVGSDADLVWNPRTKVISAKSHNLNVEYNIFEGVECRGVTVTVISQ
 RWLEEDGNLLVTPGAGRPIPRKTPDFYFKRIKARNRLAEIHGVPRLYDGPVHEVML

PAKPGSGTQARASCSGKISVPPVRNLHQSGFSLGSOADDDIARRTAQKIWAPPGGRS
 NITSLS"
 BASE COUNT 440 a 524 c 562 g 394 t
 ORIGIN

alignment_scores:
 Quality: 2803.00 Length: 572
 Ratio: 4.988 Gaps: 0
 Percent Similarity: 98.252 Percent Identity: 92.657
 alignment_block:
 US-09-367-496-8 x MMULIP4 ..

Align seg 1/1 to: MMULIP4 from: 1 to: 1920

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
 102 ATGTCTTCCAAAGCAAGAGAGCATTTCCCGGATAACGAGCGACCGCT 151
 17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
 152 TCTCATCAAAAGTGGGAAGATTGTGAACGATGACCATCTCTTCATCTG 201
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 202 ATCTGTATGTGAAGACGGTCTGATTAAACAAATTTGGAGAAATCTCATC 251
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 252 GTCCCTGGGGCATCAAAACCATCGATGCTCATGGCTGATGGTGTGCTGCC 301
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 302 TGGGGAGTGTGAGTTCACACCGCGCTGCAGATGCTGTGATGGCATGA 351
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCAGCTGATGATTCTCTCAGGCGACCAAGCGGCTCTAGCAGCGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 402 ACCACCATGATATTGGACCATGTGTTCTCTGACGCTGTGTGAGCGCTGT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 452 GGCAGCTATGAGCAGTGGCGGAGGAGCAGACAGCGCGGCTGTCTGTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 502 ACTACTCTTACATGTGGACATTCCTCGCTGCACGAGAGACCAAGAA 551
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValph 167
 552 GAGCTGGAGGCGCTAGTCAGGACAAAGGTGTGAACCTCTCTCTGCTCTT 601
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 602 CATGCATACAAAGACAGTGCACGTACTGACGGTCAGATATATGAA 651
 184 lePheSerIleAlaArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 652 TCTTCAGCGCTCCCGGACCTGGGAGCTGTGGCCAGGTGAGCGACGACAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
 702 AATGGGACATCTGGAGGAGGAGAACACAAAGCCCTGCTGGAGCAAGGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 752 CACTGGTCTCAGGGCCATGTGCTCAGCCACCCAGAGAGAGGTAGAGCGG 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250

```
|||||
802 AGGCTGTGTACAGAGCACTCACCATTGCCAAGCAGGCCAAGTCCCACTA 851
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaLeuAlaGlnAl 267
|||||
852 TACGTCACCAAGGTGATGAGCAAGGGTGCAGTGCACATGTTGCCCAAGC 901
|||||
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
902 CAAGCGCAGGGGGTGTCTCTTTGGGAACCTATCCTGCGCAGCCTGG 951
|||||
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaAla 300
|||||
952 GCACGTGATGCTCAGCTACTGAGCAAGAACTGGCGCAAGCTGCAGCC 1001
|||||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
|||||
1002 TTTGTCACTTCAACCCCTATCAACCCCGGACCTACTACTGCGAGACCACT 1051
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
1052 CACCTCTCTGTCGTCAGTGGGACCTCCAGGTGACAGGCAAGTGCCTACT 1101
|||||
334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||||
1102 GCACCTTCACTACTGCCAAGGCTGTGGCAAGACAACTTCACACTG 1151
|||||
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGl 367
|||||
1152 ATCCCCGAGGTAGTCAACGGGTATAGAAGAGGCGCATGTGTGTCTGGGA 1201
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1202 GAATGTGTGCTTCAGGGAATGGAGAGATGAGTTCGTTCGCCGGA 1251
|||||
384 hrSerThrAsnAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1252 CCAGCAAAATGCTGCCAAATCTTCAATTTTACCCAGGAGGGCGCT 1301
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1302 GTGGCCGTGGGCTGTGTGTGCTGAGCTGTGCTATCTGGAACCCAGGCGCAC 1351
|||||
417 rLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1352 GAATGTCATCTCTCCAGAGCCATAACCTGAATGTAGATACACATCT 1401
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1402 TTGAAGGAGTGGAGTGGCGAGGAGTGGCCACGGTGTATAGTCAGGGC 1451
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1452 AGAGTGTGTCTGGAGGAGGAACTGTCTGCTCACTCCAGGGGTGGCGC 1501
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1502 CTTTATTTCCCGGAGAGCTTCCCGGACTTTGTCTATAGAGGATAAGG 1551
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1552 CTCGCAACAGGCTAGCAGAGATCCAGCGTGTGCTCGAGGCTGTACGAC 1601
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1602 GGGCTGTGCTAGTATGATTTTACCTGCCAAGCCAGGAAAGTGCACACA 1651
|||||
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
|||||
1652 GGCCGTCATCTGCTTCCAGCAAGATCTCAGTGCACCCCGTGCACACC 1701
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
```

```
1702 TGCACAGTCGGGTTTCAGCCTATCTGCTCTCAGGCTGACGATCACATT 1751
|||||
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1752 GCCAGACGTACGGCTCAGAAGATCATGCACCCCCCGGAGGACGCTCCAA 1801
|||||
567 nileThrSerLeuSer 572
|||||
1802 CATCAGTCTCTTTC 1817
|||||
seq_name: gb_ro2:RNU52103
seq_documentation_block: 2489 bp mRNA ROD 02-JUL-1996
LOCUS RNU52103 Rattus norvegicus rCRMP-3 mRNA, partial cds.
DEFINITION Rattus norvegicus rCRMP-3 mRNA, partial cds.
ACCESSION U52103
VERSION U52103.1 GI:1399539
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2489)
AUTHORS Wang, L. and Strittmatter, S.M.
TITLE A family of CRMP genes is differentially expressed in the nervous
system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2489)
AUTHORS Strittmatter, S.M.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1996) Stephen M. Strittmatter, Yale
University School of Medicine, 333 Cedar Street, New Haven, CT
06520, USA
FEATURES
source 1. .2489
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
/dev_stage="adult"
<1..1696
/notes="similar to 62 kDa chick collapsin response mediator
protein"
/codon_start=2
/product="rCRMP-3"
/protein_id="AA03281.1"
/db_xref="GI:1399540"
/translation="IPRITSDRLIKGKIVNDQSFADLYVEDGLIKQIGENLIYP
GGKTTDAHGLMVLPGGVVDVHTRLQMPVLGTPADDFCQSTKAALAGGTMLDHFVP
DAGVSLAAYEQWRERADSAACDYLHVDTI PMWHESTKEELEALVRDKGVNSLVFM
AYKDRCOCTDGOIYEIIFSLRDLGAVAQVHAENGDI VVEEOKRLLEOGITGEGHVL
HPEVEAEAVYRAVTAQANCPYITKVMKGAADVAQAKRGGVYVFGPITASLG
TDGSHYWSKNWAKAAAVTSPINPDFTADHLTSLSSGDLQVTSAGHCTFTTAQKA
VGKDNFTLPIEGINGIEERMSVWKEKCVAGCKMDENEFVAVTSTNAAKIFNFPKGR
VAVGSDADLVINPRATKTVISAKSHNLNVEYNIEFEGECRGMPTVVISQGFVLEDGN
LLVTPGAGRIPTFPDFVYKRIKARNLAEIHGVPRLYDGPVHEVMLPAKPGSGT
QARASCPKISVPPVRLNHSQSGSLSGSQADDDHARRTAQKIMAPPGGRSNITLS"
```

```
BASE COUNT 557 a 692 c 708 g 532 t
ORIGIN
alignment_scores:
Quality: 2789.00 Length: 564
Ratio: 5.025 Gaps: 0
Percent Similarity: 98.404 Percent Identity: 93.262
alignment_block:
US-09-367-496-8 x RNU52103
Align seg 1/1 to: RNU52103 from: 1 to: 2489
9 llaProArgIleThrSerAspArgLeuLeuIleArgGlyGlyArgIleVa 25
|||||
```

2 ATTCCTCCGGATACGAGTACCGCCCTCTCATCAAAAGGTGGGAAGATTGT 51
 25 lAsnAspGlnSerPheTyrAlaAspValHisValGluAspGlyLeuI 42
 52 GAATGATGACCACTCTTTCATCGCGATCTCTACGTGGAGGACGGTCTCA 101
 42 leIysGlnIleGlyGluAsnLeuIleValProGlyGlyIleHisThrIle 58
 102 TTAACAAATGGAGAAATCTTATCGTCCCTGGGGAATCAAAACCATC 151
 59 AspAlaHisGlyLeuMetValLeuProGlyGlyValAlaSpValHisThrAr 75
 152 GATGCTCATGGTCTGATGCTGCTGGGGAGTTGACGTCCACACCCG 201
 75 gLeuGlnMetProValLeuGlyMetThrProAlaAspPheCysGlnG 92
 202 GCTGACGATGCTGCTGGCATGACCCAGCCGCGATGATTTCTGTCA 251
 92 lYThrLysAlaAlaLeuAlaGlyIleThrThrMetIleLeuAspHisVal 108
 252 GCACCAAGGCTGCCCTAGCAGCGGGACCAACGATGATATTGACCATGTG 301
 109 PheProAspThrGlyValSerLeuLeuAlaAlaTyrGluGlnTrpArgG 125
 302 TTTCTGATGCTGGCGTGAGCTGCTGGCGGCTATGACGAGTGGCGGA 351
 125 uArgAlaAspSerAlaAlaCysCysAspTyrSerLeuHisValAspIleT 142
 352 GCGAGCAGACAGTGCAGCGCTGTGTGACTACTCTTACACGTGGACATTC 401
 142 hArgTrpHisGluSerIleLysGluGluLeuGluAlaLeuValLysGlu 158
 402 CCGCGTGGCATGAGACACCAAGAGGAGCTGGAGGCGCTTAGTAGGAC 451
 159 LysGlyValAsnSerPheLeuValPheMetAlaTyrLysAspArgCysG 175
 452 AAAGTGTGAATCTCTCTGCTGCTTCAAGCTTACAGGACAGGTGCCA 501
 175 nCysSerAspSerGlnMetTyrGluIlePheSerIleIleArgAspLeuG 192
 502 GTGTACTGATGGCCAGATATATGAATCTTCAGCCTCATCCGGGACCTGG 551
 192 lYAlaLeuAlaGlnValHisAlaGluAsnGlyAspIleValGluGlu 208
 552 GAGCTGTGGCCAGGTGATGAGAAAACGGGACATCTGTGGAGGAGAA 601
 209 GlnLysArgLeuLeuGluLeuGlyIleThrGlyProGluGlyHisValIle 225
 602 CAGAAAGCGCTCTGGACAGGCGATCACTGGTCTGAGGGCCATGTGCT 651
 225 uSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThrI 242
 652 CAGCACCCAGAAGAGGTAGAGGCGGAAGCTGTGTACAGAGCTGTGACCA 701
 242 leAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerLys 258
 702 TTGCCAACAGCCACNCTGCCCATCTGTATATACCAAGGTGATGAGCAAG 751
 259 GlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValph 275
 752 GGTGCGGTGATATGTTGTTGCCAAGCCAGCGCAGGGGGGTGGTCTCT 801
 275 eGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrps 292
 802 TGGGGAACCTATCACTGCCACCTGGGACCTGATGGCTCACACTACTGGA 851
 292 erLysAsnTrpAlaLysAlaAlaPheValThrSerProProValAsn 308
 852 GCAGAACTGGGCCAAGGCTGCACCTTCGTACATCACCCCTTATCAAC 901
 309 ProAspProThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAs 325
 902 CCGGACCTACTACTGACGACCACTCACCTCTCTGCTGTCTCCAGTGGGA 951

325 pLeuGlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnLys 342
 952 CCTCCAGGTGACAGGAGTGCACACTGCACCTTCACTACTGCCCAAGG 1001
 342 laValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIle 358
 1002 CTGTTGGCAAGATAACTTCAAGCTGATCCCGAGGCGCATCAACGCTATA 1051
 359 GluGluArgMetSerMetValTrpGluLysCysValAlaSerGlyLysMe 375
 1052 GAAGAACGCGATCTGTGTGCTGGGAAATGTGTGGCTTCAGGGAAT 1101
 375 tAspGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIleP 392
 1102 GCACGAGATGAGTTCGTTGCCGTGACACGACCAAAATGCTGCCAAATCT 1151
 392 hAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAsp 408
 1152 TCAATTTCTACCCAGGAAGGCGGTGTGGCTGTGGCTCTGATGCTGAC 1201
 409 LeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysThrHi 425
 1202 CTGGTCACTCTGGAACCCAGGCGCACAAAGTCATCTCTGCCAAGACCA 1251
 425 sAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyA 442
 1252 TAACTGTAATGTAGAGTACACATCTTCGAGGAGTGGAGTGCAGAGGCA 1301
 442 laProAlaValValIleSerGlnGlyArgValAlaLeuGluAspGlyLys 458
 1302 TGCCACGCTGCTCATAGTACAGGCGAGAGTGTGTGGAGGACGGAAC 1351
 459 MetPheValThrProGlyAlaGlyArgPheValProArgLysThrPhePr 475
 1352 CTGCTGTGCTACCTCCAGGGGCTGGCGCTTCATTCCCGGAAGACATTCC 1401
 475 oAspPheValTyrLysArgIleLysAlaAlaArgAsnArgLeuAlaGluIleH 492
 1402 AGACTTGTCTATAGAGGATAAAGGCTCTGAACAGGCTAGCAGAGATCC 1451
 492 lGlyValProArgGlyLeuTyrAspGlyProValHisGluValMetVal 508
 1452 AGGCTGTGCCCGAGCGCTGTATGATGGCGCTGTCCAGAGTGATGTTA 1501
 509 ProAlaLysProGlySerGlyAlaProAlaArgAlaSerCysProGlyLy 525
 1502 CTGCCAAGCCAGGAGCGGCACACAAAGCCGTGCTCCAGAGTCCAGGCAA 1551
 525 sIleSerValProValArgAsnLeuHisGlnSerGlyPheSerLeus 542
 1552 GATCTCAGTGCACCCGTGCGCAACTGCACCACTCAGGTTTCAGCCTAT 1601
 542 erGlySerGlnAlaAspAspHisIleAlaArgArgThrAlaGlnLysIle 558
 1602 CTGGCTCTCAGGCTGACGATCAGATTGCCAGACGACAGCTCAGAAGATC 1651
 559 MetAlaProProGlyGlyArgSerAsnIleThrSerLeuSer 572
 1652 ATGGCACCCCGGAGGAGCGCTCCCAACATCACTTCACTTCTTCT 1693

seq_name: gb_ov: AF249294

seq_documentation_block: 2470 bp mRNA VRT 06-DEC-2000
 LOCUS AF249294
 DEFINITION Gallus gallus collapsin response mediator protein 3 (CRMP3) mRNA,
 complete cds.
 ACCESSION AF249294
 VERSION AF249294.1 GI:11559529
 KEYWORDS chicken.
 SOURCE Gallus gallus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

REFERENCE
AUTHORS      Fukuda,M., Wataabe,I., Yuasa-Kawada,J., Kawachi,H., Kurolwa,A.,
TITLE        Molecular Characterization of CRMP5, a Novel Member of the
              Molecular Response Mediator Protein Family
JOURNAL      J. Biol. Chem. 275 (48), 37957-37965 (2000)
PUBMED       10956643
REFERENCE     2 (bases 1 to 2470)
              Yuasa-Kawada,J., Wataabe,I. and Noda,M.
              Direct Submission
              Submitted (27-MAR-2000) Division of Molecular Neurobiology,
              National Institute for Basic Biology, Myodaiji-cho, Okazaki, Aichi
              444-8585, Japan
FEATURES
source       Location/Qualifiers
              1..2470
              /organism="Gallus gallus"
              /db_xref="taxon:9031"
              /tissue_type="retina"
              /dev_stage="embryonic day 8"
              /note="breed: White Leghorn"
              1..2470
              /gene="CRMP3"
              14..1738
              /gene="CRMP3"
              /codon_start=1
              /product="collapsin response mediator protein 3"
              /protein_id="AAG37997.1"
              /db_xref="GI:11559530"
              /translation="MSPNKRSGRGTSEQLIKGKIIVNDQSFYDIYVEDGLIKQ
              IGENLVPSGVRTVDAYGOLVPGGIDVHTRLQMAVMGASADDFYOGTRALAGGT
              MIMDLVPEAGTSLIAAYEOWRQADSRACDYALHIDIPRWHESELEALVKDKG
              VNSFLVWAYKDLQCTDAOMYEIFCIIRDLGALQAQVHAENGDIIEEQRKLLDIGIT
              GPGEHLVSRPEVEAEVYRAITIAKQANCLPYVTKIMSKSAADVVAQAKRKGTVYTG
              EPTASLGADSHVSKWAKAAAFVTPSPSPDPTPERLSLLSCDLQVAGSAHC
              TPTTAOKAGKNDTLDIPEGTNGIEERMAIIWEKCVYSGKMDNDFVAVTSTNAKIF
              NCYPRKRIAGVSDADLVLMNPRAFKVTSKTHNLNVEYNIFEGTECHGAPAVVISOG
              KVLBDGNLFYTESGSEVPRKTPDFYVRIKARNRLAEVHGVPRLGIDGPFVDVIL
              STKAVPTTAAIRIAACAGKVPAPVPRNLHQSGFSLSGSQADHDHVARRTAQIMAPPGG
              RSNITSLs"
BASE COUNT   533 a 685 c 757 g 495 t
ORIGIN
alignment_scores:
  Quality: 2517.00      Length: 574
  Ratio: 4.618          Gaps: 1
  Percent Similarity: 94.948 Percent Identity: 81.359
alignment_block:
  US-09-367-496-8 x AF249294 ..
Align seg 1/1 to: AF249294 from: 1 to: 2470
1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
|||||
14 ATGTGGAACCCGGCAGAGAGCGGCCCGCGGACGAGCGAGCAGCT 63
|||||
17 uLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
|||||
64 GCTGATCAAGGGGGGAAATCGTCAGCAGCAGCAGCTCGTTTATGCG 113
|||||
34 spValHisValcluaspglyLeuIleLysGlnIleGlyGluAsnLeu 50
|||||
114 ACATTTATGTCGAAGATGGTGTGATAAACAGATTGGAGAACCTGGCT 163
|||||
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeu 67
|||||
164 GTTCCACGGGGGTGAGGAGGTGTGATGCCACGGGACGCTGGTGCC 213
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
|||||

```

```

214 GGGGGGCGATCGATGTCACACGGCGCTGCAGATGGCTGTGTGATGGGATGG 263
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||
264 CCTCTGTGTGATGACTTCTACCAAGGCACGAGGCGGCTGGCGGGAGGA 313
|||
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
|||
314 ACCACGATGATAATAGGACCATGCTGCTGCCGAGGCTGGGAGGCTGCT 363
|||
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
|||
364 GGCCTGCTACGAGCAGTGGCGGACGCTGCTGACAGCAGGCGCTGCTGCG 413
|||
134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
|||
414 ACTAGCGCTGCATGCATGCCCGCTGGCATGAGAGCCTGCGGGAG 463
|||
151 GluLeuGluAlaLeuValLysGlyLysGlyValAsnSerPheLeuValPh 167
|||
464 GAGCTGGAGGCGCTGGTCAAGACACAGGCTGTGAACCTCTTCTTGGTTT 513
|||
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||
514 TATGCGCTACAAAGGACAGCTGCTGACGACGCGCCAGCATGTATGAA 563
|||
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||
564 TATTCTGCTATTATCGAGACCTGGGGGCTATAGCCCAAGTGCACGCTGAA 613
|||
201 AsnGlyAspIleValcluaspglyGlnLysArgLeuLeuGluLeuGly 217
|||
614 AACGGGACATCATCGAGGAGCAGCAGAGAGGCTGCTGGACATTTGGAT 663
|||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
|||
664 CACTGGGCGAGAGGCGCAGCTGCTGAGCGGCTGAGGAGGTGAAGCAG 713
|||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||
714 AGCGGTGTACCTGCAATACATAGCAAAACAGCCCAACTGCCCTTG 763
|||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
|||
764 TAGCTCACCACCAATAATGAGCAAAAGTGCAGCAGATGTGGTAGCTCAAG 813
|||
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
|||
814 AAAGAAAGAGGACCTGTGTGTGTCAGGAGGAGCCCATCATCTGCCAGCT 863
|||
284 lYThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||
864 GTGCTGACGCTCGCACTACTGAGCAAGAACTGGGCAAGGCTGCAGCC 913
|||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHis 317
|||
914 TTGCTCACCTCACCCCGCTCAGCCCGGACCCACCCAGCCAGAGCGCCT 963
|||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHis 334
|||
964 CTGCTCCCTGCTGCTGCTGGGACCTGTCAGGTGGCAGGAGTGTCTACT 1013
|||
334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
|||
1014 GCAGTTTCACTACGCGCCAGAAAGCGCTGGGAAGGACAACTTCACCTC 1063
|||
351 lIleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 367
|||
1064 ATCCCGGAGGCGGAGATGTCATGCGAGGCGGATGGCCATAATCTGGGA 1113
|||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
|||
1114 GAAGTGGCTGTCTCAGGGAAGATGGATGAGACGATTTCTGTCGGGTGA 1163

```

384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1164 CCAGCACTAATGCTGCAAAATCTTTAACTGCTACCCAGGAGGGGA 1213
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 417
 1214 ATCCGGGGTGGCTGATGATGAGTGTGCTGTGTGGAACCCAGGCTAC 1263
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTrpAsnIleP 434
 1264 AAAAGTCATCTCGGCAAAACCCACAAATTTGAATGTGAGTACACATAT 1313
 434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
 1314 TTGAAGGACGAGGTGTACGCGGCTCTGCTGTGTGTCATCAGTCAGGCG 1363
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1364 AAAGTCGTCTTGAGGACGGGAACCTGTTGTCCACCGAGGCTCGGGCG 1413
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 1414 CTTCGTTCCCAAGAAAGAGCTTCCCGGATTTGTTTATAAAGAATAAGG 1463
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1464 CGAGGAACAGGCTGGCGGAGGTGCACGGTGTCTCCACCGGGGCTGTATGAC 1513
 501 GlyProValHisGluValMetValProAlaLys.....ProGlySerGI 515
 1514 GGCCGGTCCACGATGTCGTCCTCAGCACCAAGCTGTGCCACCATATCGC 1563
 515 yAlaProAlaArgAlaSerCysProGlyLysIleSerValProValA 532
 1564 GGCTCCAGGATCGCGCTGCGCGGCAAGTCCCGCGTCCCGTGC 1613
 532 rGAsnLeuHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAsp 548
 1614 GCAACCTGCACAGTGGGGTTCAGCTGTCCAGGCTCTCAGGCTGACGAC 1663
 549 HisIleAlaArgArgThrAlaGlnLysIleMetAlaProProGlyLys 565
 1664 CAGTTGCGCGGCGCACAGCCAGAGATCATGGCCCCCGCGCGCG 1713
 565 gSerAsnIleThrSerLeuSer 572
 1714 CTCACATCACCCTCCTGTGCC 1735

seq_name: gb_ov:GGU17277

seq_documentation_block: 2423 bp mRNA VRT 02-APR-1996
 LOCUS GGU17277
 DEFINITION Gallus gallus collapsein response mediator protein CRMP-62 mRNA,
 complete cds.
 ACCESSION U17277
 VERSION U17277
 KEYWORDS U17277.1 GI:882146
 SOURCE chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 2423)
 Goshima,Y., Nakamura,F., Strittmatter,P. and Strittmatter,S.M.
 Collapsein-induced growth cone collapse mediated by an intracellular
 protein related to UNC-33
 Nature 376 (6540), 509-514 (1995)
 95364923
 2 (bases 1 to 2423)
 Strittmatter,S.M.
 Strittmatter,S.M.
 Direct Submission
 Submitted (17-NOV-1994) Stephen M. Strittmatter, Neurology, Yale

University School of Medicine, 333 Cedar Street, New Haven, CT
 06520, USA
 FEATURES
 source
 1. 2423
 Location/Qualifiers
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="dorsal root ganglion"
 /dev_stage="embryonic day 7"
 302..2020
 CDS
 /note="62 kDa collapsein response mediator protein"
 /codon_start=1
 /product="CRMP-62"
 /protein_id="AAA93200.1"
 /db_xref="GI:882147"

alignment_scores:
 Quality: 2362.00 Length: 571
 Ratio: 4.457 Gaps: 0
 Percent Similarity: 92.820 Percent Identity: 76.357
 alignment_block:
 US-09-367-496-8 x GGU17277 ..
 Align seg 1/1 to: GGU17277 from: 1 to: 2423

BASE COUNT 593 a 613 c 654 g 563 t
 ORIGIN

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 302 ATGCTTTACCGGGGAGAAAGAACATCCCGCGATCACGACGATCGTCT 351
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
 352 TCTGATCAAGGTGGGAGATCGTCAATGACGACCATCTTCTATGCAG 401
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 402 ACATCTACATGGAAGATGGGCTCATCAAGCAATAGGGAGATCTCATC 451
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 452 GTGCCCGAGGAGTGAACACCATCGAAGCCCGCAGGAGGATGATGTC 501
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84.
 502 TGGAGGGATCGATGTTCAACCGCTTCCAGATGCCAGACGAGGGATGA 551
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 552 CTCTCGTATGATCTCTTCCAAAGGAGCAAGGCTCGCTGCTGGGGC 601
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 602 ACCACCATGATCATGATCATGTGTTCTCTGAGCCCGGACCATGTTGT 651
 117 uAlaAlaTyrGluGlnTrpArgGluAlaAspSerAlaAlaCysCysA 134
 652 GACTGCATTTGACCATGAGAGATGGCAGACACAAATCGTCTGTG 701
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 |||||

Tue Jul 31 13:08:18 2001

```
702 ATTACTCCTCGCAGCTGGACATCCAGGAGTGCACAAAGGGCTCCAGGAG 751
151 GluLeuGluAlaLeuValValGluLysGlyValAlaAsnSerPheLeuValph 167
752 GAGATGGAAGCTCTGGTTAAAGATCAGCGTGGAATTCATCTCCTGGTTA 801
167 eMetaLAtyLyAspAtqCysGlnCysSerAspSerGlnMetTyrGluI 184
802 TATGGCTTCAAGACCGCTTCCAGCTGCTGCAGCTCGCAGATCTATGAGG 851
184 lePheSerlleleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
852 TCCTGAGTGTATCCGGGATATGTTGGCCACAGCTCAAGTGCATGCTGAG 901
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
902 ATGTGTGACATCATTCCTCCGAGGAGCAGCAAGATCTCTGGAGTGGAT 951
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
952 CACAGCCCTGAAGGCGACGTGTTGAGCAGACCTGAAGAGGTGGAGGCG 1001
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
1002 AAGCTGTGAACCGGGCAATCACCATCGCCACCAACCAACTGCCCTCTC 1051
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
1052 TATATCACCAAGGTGATGAGCAAAAGCGTGTGAAGTCATTCCTCAAGC 1101
267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
1102 CAGGAAAAGGCGACCTGTGTGTATGGGAGGCCCATCACAGCAGCTGG 1151
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
1152 GTACTGATGATCTCATCTGAGCAGCAGATGTTGGGGAAGGCGCGCT 1201
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
1202 TTTGTCACTTCCCAACCACTGAGTGTGCTGACCAACCACTCTGATTTCT 1251
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
1252 CAACCTCAACACTGCTGCTGCTGCGGGACCTCCAAAGTTACTGGCAGTCC 1301
334 ysThrPheThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
1302 CCACCTTCAACACTGCTCAGAAAGCTGTGGGAAGGACAACCTTCAACCTA 1351
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTyrP 367
1352 ATCCCGAGGGGACCAACGGGACTGAGGAGAGGATGTCATCATCTGGGA 1401
367 uLysCysValAlaAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
1402 CAAGCTCTGTGTGATGCTGCAAGATGATGACAGACCACTGTTGGCTGGA 1451
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgGlyGlyArg 400
1452 CCAGCACAATGTCAGCTAAATCTTTAACTTGTATCCCTGATCCACGGA 1501
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
1502 ATTCAGTGGGCTGCTGATGCTGATTTGTTATCTGGGATCTGCAGCGGT 1551
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1552 CAAGACAATCTCTGCAAGACCCCAACATATCTCTGGAGTACAACATCT 1601
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1602 TTGAAGGCATGAGTGTGCTGGTCTCTCCCTGTTATCATCAGCAAGG 1651
```

451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
1652 AAGATCGTGTGGAGGATGGAAATCTCCACGTAAAGGAGGGTCTGGAG 1701
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1702 GTAATATCCCAAGAACCATTCCTGACTGCTGTTTACAAGGCATCAAG 1751
484 laaAgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1752 CAAGGAGCAGGCTGCGGAGCTGCGGGTGTGCTCGAGGTTCTCTATGAT 1801
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1802 GGGCGGCTCTCGAAGTGTGCTGACGCCGAGAGCGGTCACTCTCGCATC 1851
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
1852 TTCAGCTAAACATCTCTGCCAAACAGCAAGCCACCTGTGAGGAACC 1901
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
1902 TGCACACGCTCTGGCTTACGTTGTCTGGCGCACATGATCGACGACCAATC 1951
551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
1952 CCACGCGGCGACCCAGCGCATCTGCGCACCGCGGCGGACGTGCCAA 2001
567 nIleThrSerLeu 571
2002 CATCACGAGCTTG 2014

seq_name: gt_om:BTU83278

seq_documentation_block: 3884 bp DNA MAM 08-OCT-1997
LOCUS BTU83278
DEFINITION Bos taurus neural specific protein CRMP-2 gene, complete cds.
ACCESSION U83278
VERSION U83278.1 GI:1916226
KEYWORDS
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3884)
AUTHORS Kamata,T.K.
TITLE Isolation and characterization of a neural specific protein (Nsp60)
gene homologous to C. elegans unc-33 controlling axonal outgrowth
and guidance
JOURNAL Brain Res. Mol. Brain Res. (1997) In press
REFERENCE 2 (bases 1 to 3884)
AUTHORS Kamata,T.K.
TITLE Direct Submission
JOURNAL Submitted (27-DEC-1996) Lab. of Biochemical Physiology, National
Cancer Inst. FCRC, P.O. Box B, Bldg 567, Frederick, MD 21703, USA
FEATURES
SOURCE
1..3884
/organism="Bos taurus"
/db_xref="taxon:9913"
/note="map location 8p21 in human; mouse 14"
177..1895
/note="unc-33 homolog"
/codon_start=1
/product="neural specific protein CRMP-2"
/protein_id="AAB80618.1"
/db_xref="GI:1916227"
/translation="MSYQSKKNIPRTSDRLIKGKIVNDQSFVADITYMEDGLIKQ
IGENLIVPGVKTIHAHSRWIPGGIDVHTRFQMPDQGTADDFFQGTKAALAGSTT
MIIDRVYEPGTSLLAAFDQWREWDKSCDDSLRVDITEWHKGVQEMEARLYDGH
VNSFLVYMAFKDRFQTSQIYEVLSVIRDIGAIVHAENGDIIEAQORILDIGIT
GPEGHVLSPREEVEAEAVNKSITIANQTNCLPYITKVMSSAAEIVIAQARKKGTYYVG

EPITASLIDGGSHYKKNWAKAAAFVTPPLSPDPPTPDFLNSLLSCDILQVTSACAH
TFTNAQKAVGKONFTLLPGCTNGTTERMSVWDRAVITGKMDENQFVAVTSTNAKYF
NLIPRKGRAGVSGDGLVDPDSVKTSAKTHNSLBYETFCGECRGSPLVVISOG
KIVLEDTLHVTEGSGYIPRKPFDFVYKRIKARSALAEURGPRGLDGPDPVCSG
TPKTVTPASSAKTSPAKQAPVPRNLHGSGSLGAGIDNNIPRTRQIVAPGGR

BASE COUNT	NITS LG*		
ORIGIN	911 a	1031 c	912 t
		1029 g	1029 g
			1 others

```
alignment_scores:
  Quality: 2347.00
  Ratio: 4.437
  Percent Similarity: 92.644
  Length: 571
  Gaps: 0
  Percent Identity: 75.657
```

alignment_block:
US-09-367-496-8 x BTU83278

Align seg 1/1 to: BTU83278 from: 1 to: 3884

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
177 ATGTCCTTATCAGGGGAGAAAAATATTCCAGGCATACAGACTGATCGTCT 22

17 uLeulleArgCylGlyArgIleValAsnAspAspGlnSerPheTyrIaa 34
| | | | : | | | : | | | : | | | : | | | : | | | :
227 TCATCATCAAGGAGGTAAATGGTTAATGATGACCACTCTTCCTATGCAG 278

34 spValHisvalGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
|||:::|||||
277 ACATATACATGGGAAGATGGGTGATCAAGCAAATAGGAGAAAACCTGATT 328

51 ValProGlyCysIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 |||||
 327 GTCCCGGGAGGGGTGAAGCATTTGAAGCCATTACAGAATCGTCGATCC 378

67 oGlyglyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
|||||:::|||||||
377 TGGAGGGATTGACGTCCATACCCTTCACAGATCCTCATACGGCAATCA

84 hrProLaaspasppheCysGInglyThrllysAlaAlaIeuAlaGlyGly 1000
|| |||||
427 CCTCTGCTGATCACTTCTTTTCAGGCACCAAGSCAGCCCTCCCCCGC 476

101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe
477 ACCACATCATCTTCCACCACCTCCTGGTTTCCTCGAGGAGGA

117 uAaLaATyTrGluInTrpArgGluArgAlaAspSerAlaLaCysCysA 134
144 ..n..n..n..n..h..i..c..a..l..o..n..c..c..a..c..g..g..g..t..t..c..t..t..g..a..c..c..t..t..g..g..g..a..c..a..a..g..o..c..t..t..g..c..t.. 152

[illegible][illegible][illegible]

0 // CATGGCTTTTCAAAGATCGTTTTCCAGCTCACGGATTTCCCGAGATCTATGAAG 726

121 TCCTGAGCGTGATCCGGGATATTGGAGCGATTGCCCCAGGTCCACCGGAA 776

201 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuGluLeuGlyIyl 217

777

777 AACGGGACATCATCGCAGAGGACGACAGAGGATCCTGGATCTGGGCAT 826

217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234

|||||
827 CAGGGCCCTGAGGACATGTCTCTGAGCGGCCTCAGGAGGTTGAGGCTG 876

|||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCyProLeu 250
|||||
877 AAGCGTGAAACGTTCCATCACCATCGCCACCAGACCACTGCCCTCTG 926

|||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
927 TATCATCACCAAGGTGATGAGAAGAGCGCGGAAGTTATCGCCAGGC 976

|||
267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
977 CCGAAAAGGAACCGTGGTGTCAGCGAGCCCATCACCGCCAGCTGG 1026

|||
284 lyThrAspGlySerHisTyrTrpSerLysAsnTpAlaLysAlaAlaA 300
1027 GGACCGAGGGCTCTCACTACTGGAGCAAGAACTGGGCCAAGGCTGTGCT 1076

|||
301 PheValThrSerProProValAsnProAspproThrThrAlaaspHisLe 317
1077 TTGTGTCACTCCCGCGCCCTGAGCGCCGACCAACCACTCCAGCACTTCT 1126

|||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
1127 GAATCTCTGCTGCTGTGGAGATCTCCAGGTGACGGCAGTGCTCACT 1176

|||
334 ystThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
1177 GTACTTCAACACGGCCCGAAGAGCGGTGGAAAGACAACAACTTCACGCTG 1226

|||
351 lleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValtrpGl 367
1227 ATCCTGAGGSCAACACGGCAGGAGCGGATGCTGTCACTCTGGGA 1276

|||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
1277 TAAGCTGTGGTCTACTCGGAAGATGACAGAACCAAGTTGTGGCCGTGA 1326

|||
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
1327 CCASCACCAATGTCGCAAGTCTTCACTGTACCCCGCGAAGGCGCGC 1376

|||
401 ValAlaValGlySerAspAlaAspLeuValIleTpAsnProLysAlaIth 417
1377 ATTGCTGTGGGTCTATGCTGACTTGGTCACTCGSACCCCGACAGCGT 1426

|||
417 rLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
1427 GAAGACATCTCGCAAGACCCACACAGCTCTCTTGAGTACAACATCT 1476

|||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
1477 TCGAAGGCATGGAGTCGCGAGGCTCCCCACTGGTGTCTATCAGCCAGGG 1526

|||
451 ArgValAlaLeuLuspgLyllysMetPheValThrProGlyAlaGlyAr 467
1527 AGATTGTCTGGAGATGGTACCCCTCATGTCACCGAAGGCTCTGGCG 1576

|||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
1577 CTACATCCCGCGAAGCCCTTCCCCTGACTTCGTTTACAAGCGTATCAAG 1626

|||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
1627 CAAGGACAGGCTGGCGAGCTCGAGAGGGGTTCGCCGTGGTCTGTAGCAC 1676

|||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
1677 GGACCTGTGTGAGGTGTCGTGTACACCCAGACGGTCACTCCGCGCTC 1726

|||
517 olaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
:::

1727 CTGGCTAAGACGCTCTCCGGCCAGCAGCAGGCCACCTGTCGGGACCC 1776
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
 1777 TGCACCACTGCTGATTCAGCTTCTCCGGTCCCGAGATTGATGACAATC 1826
 551 AlaArgArgThrAlaGlnLysIleMetAlaProGlyGlyArgSerAs 567
 1827 CCTCGCCGACACACGCGCATCGTGGCGCCCTGGTGGCGCGTGTAA 1876

567 nileThrSerLeu 571
 1877 CATCACCAGCGCTG 1889

seq_name: gb_ro2:RRTOAD64

seq_documentation_block: 2947 bp mRNA ROD 23-JAN-1996
 LOCUS RRTOAD64
 DEFINITION R.rattus mRNA for T0AD-64.
 ACCESSION Z46882
 VERSION Z46882.1 GI:599965
 KEYWORDS TOAD-64.
 SOURCE black rat.
 ORGANISM Rattus rattus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 2947)
 AUTHORS Minturn,J.E., Fryer,H.J., Geschwind,D.H. and Hockfield,S.
 TITLE TOAD-64, a gene expressed early in neuronal differentiation in the
 rat, is related to unc-33, a C. elegans gene involved in axon
 outgrowth

J. Neurosci. 15 (10), 6757-6766 (1995)

69033765

REFERENCE 2 (bases 1 to 2947)

Hockfield,S.

AUTHORS Direct Submission

TITLE Submitted (14-NOV-1994) Susan Hockfield, Neurobiology, Yale

JOURNAL University School of Medicine, 333 Cedar Street, New Haven, CT,
 06510, USA

FEATURES Location/Qualifiers

source

1..2947
 /organism="Rattus rattus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10117"
 /clone="TOAD64 (turned on after 'division, 64kd)"
 /tissue_type="brain"
 /clone_lib="P5 rat pBluescript brain library"
 178..1896

CDS

/codon_start=1
 /product="TOAD-64"
 /protein_id="CA886981.1"
 /db_xref="GI:599966"
 /db_xref="SWISS-PROT:P47942"
 /translation="MSYQKGNIPRITSLRLIKGGKIVNDQSFVADITYMEDGLIKQ
 IGNLIVPGVKVITIAHSRMVTPGGIDVTRFQMPDQMTSADDFQGTKAALAGETT
 MTDHVPVGGVSLIAFDQWREWADSKCCDYSLHVDITENHKGTOERMEALVKDGH
 VNSFLVMAFKRQITDSQIYEVLSVIRDIGALQVHAENGDIATFEOORILDGIT
 GPBGHVLSPREEAEAVNRKASNTIANTCPLFTVKVMSKRAEVTQARKKGTVVYG
 EPTASLSDGSHYWSKNWAKAAAFVTSPLSPDPTTDFLNSLSLSCDLQVTGSAHC
 TFNTAQKAVGKDNFTLIPETNGTEERMSVINDKAVYTKMDENQFVAVTSTNAKVF
 NLYPRKGRISVSDADLVTDPSVKTISAKTHNSALEYNIPEGMECRGSPLVVISQ
 KIVLEDGTLHVTGSGRYIPRKPFDPFVKIKARSLAEIRGVPRGLYDGPVCEYSV
 TPKTVTPASSAKTSPAKQAPPVNRNLHQSFSLSGAIDIDNIPRRTQRIVAPPGGRA
 NITSLSG"

BASE COUNT 666 a 765 c 769 g 747 t

ORIGIN

alignment_scores:

Quality: 2343.00 Length: 571

Ratio: 4.438 Gaps: 0

Percent Similarity: 92.469 Percent Identity: 75.832

alignment_block:

US-09-367-496-8 x RRTOAD64

Align seg 1/1 to: RRTOAD64 from: 1 to: 2947

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 178 ATCTCTTATCAGGGAAGAAAATATTCACGCATCACCAGCGATCGTCT 227
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
 228 TCTGATCAAAAGGTGGCAAGATTGTGAATGATGACCATCTCTTATGCAG 277
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 278 ACATATACATGGAAGATGGGTGTCAAGCAATAGGAGAAAACCTGATT 327
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 328 GTCCAGGAGGGGTGAAGACCATCGAAGCCCACTCCAGAAATGGTGATCCC 377
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
 378 TGAAGGAATTGACGTGCACACTCGCTTCCAGATGCCAGACCAAGGAATGA 427
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 428 CATCAGCTGATGACTTCTTCCAGGGAACCAAGCAGCCCTGGCCGAGGA 477
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 478 ACCACCATGATCATCGACCATGTTGTTCTGAGCCCGGACCAAGCCTATT 527
 117 uAaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 528 GGCAGAGCTTTGATCAGTGGAGGAGTGGCGGACAGCAAGTCTCTGTGTG 577
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 578 ACATATCGCTGCAGTGGACATCAGGAGTGGCACAAGGCGCATCCAGGAG 627
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 628 GAGATGGAAGCTCTGCTGAAGGACCAAGCGGTAAACTCTCTCTCGTSTA 677
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 678 CATGGCTTTCAAAGATCGTTCCAGCTGCGGATTTCCAGATCTATGAAG 727
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 728 TACTGACCGTGTATCCGGGATATTTGTCCTCATAGTCAAGTCCATGCAGAG 777
 201 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuLeuGlyIle 217
 778 AATGCTGACATCATTTGACAGAGAAACAGCAGAGGATCTGATCTGGGCAT 827
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 828 CACAGGCCCCGAGGAGACGCTGCTGACCGCCGACAGAGGATCGAGGCTG 877
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 878 AA3CTGTGAACCGGTCTCATCCATTGCCAATCAGACCAACTGCCCGCTG 927
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 928 TATGTCAACCAAGGTGATGACCAAGAGTGTCTGCTGAAGTCAATCGCCAGGC 977
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
 978 ACGGAAGAGGGAACACTGTGGTGTATGTGTGAGCCCATCACTCCAGCCTGG 1027

```
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
|||||
1028 GGACTGATGGCTCTCATATTGGAGCAAGAACTGGCCAGGCGCTGCC 1077
|||||
301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
|||||
1078 TTTGTACCTCTCCACCTTGGACCTTGGAGCCGACCAACCACTCCAGACTTCT 1127
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGluValThrGlySerAlaHisC 334
|||||
1128 CAACCTGCTGCTGCTGCTGGAGCACTCCAGGTCACGTGAGGAGTCCCACT 1177
|||||
334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPhAlaLeu 350
|||||
1178 GTACCTTCAACACTGCCAGAGGCTGTGGGAAGGATAACTTCCACCTG 1227
|||||
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGI 367
|||||
1228 ATCCAGAGGGCACCACATGACCTGAGGAGCGGATGCTGTCTATTGGGA 1277
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1278 TAAAGCTGTGCTCACTGGGAAGATGGACGAGAACCACTTTGTGCTGTA 1327
|||||
384 hrSerThrAsnAlaAlaLysIlePheAsnPhetYrProArgLysGlyArg 400
|||||
1328 CTAGCACCAACGAGGCAAGTCTTCAATCTTACCCACGAGAAAGTCTGT 1377
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1378 ATCTCGTGGATCTGACGACGACCTGGTGATCTGGGACCTGACAGTGT 1427
|||||
417 rLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1428 GAAGACCATCTGTGCAAGACGACCAACAGTCTCTTGTAGTACACATCT 1477
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1478 TTGAAGCATGAGTGTGGGGTCTCCCACTGGTGTCTATCAGCCAGGCG 1527
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1528 AAGATTGTCTGGAGGAGCGACGTTGCATGTACAGGAGGCTCAGACG 1577
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1578 CTACATTCCTCCGAGGCTTCCCTGACTTGTGTACAAACGATCAAGG 1627
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1628 CAAGGAGCAGCTGGCTGAGCTGAGGGGGTCCCTGCTGGCCTGTATGAT 1677
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1678 GGACCGGTATGCGAGGTGCTGTGAGCCCAAGAGCGGTCACTCCGCGCTC 1727
|||||
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
|||||
1728 ATCAGCTAAGACATCCCTGCGCAAGCAGCAGGCGCACCTGTTCGGAACC 1777
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1778 TGCACCATGCTGTTTACCTTGTCTGTGCTGCTGATGACGACCAACAT 1827
|||||
551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1828 CCCCAGCCGACCCAGCAGCATTTGTGGCCGCCCTGCTGGCCGCGCA 1877
|||||
567 nIleThrSerLeu 571
|||||
1878 CATCAGGCGCTG 1890
```

seq_name: gb_pri0:HSU17279

seq_documentation_block: 1829 bp mRNA PRI 02-APR-1996

LOCUS HSU17279

DEFINITION Human collapsin response mediator protein hCRMP-2 mRNA, complete cds.

ACCESSION U17279

VERSION U17279.1 GI:1244399

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1829)

AUTHORS Goshima, Y., Nakamura, F., Strittmatter, P. and Strittmatter, S. M.

TITLE Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33

JOURNAL Nature 376 (6540), 509-514 (1995)

MEDLINE 95364923

REFERENCE 2 (bases 1 to 1829)

AUTHORS Strittmatter, S. M.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-1994) Stephen M. Strittmatter, Neurology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA

COMMENT On Apr 2, 1996 this sequence version replaced gi:882150.

FEATURES

Location/Qualifiers

1..1829

/organism="Homo sapiens"

/db_xref="taxon:9606"

72..1790

/note="collapsin response mediator protein"

/codon_start=1

/product="hCRMP-2"

/protein_id="AAA93202.1"

/db_xref="GI:1244400"

/translation="MSYQGKKNIPRTSDRLIKGKIVNDQSFYADLYMEDGLIKO IGENLIVPGVKTIEARSMVIPGGIDVHTRFQMPDQGMTSADDFQCTKAALAGTTF MIDHVPPEPGTSLLAARMDREWADSKSCDYSIHVDI SEWHKGIQEMEAALYKDHG VNSFLVYMAFKDRLQDQIYEVLSVRDICAQVHAENGDIIEEQQRILDIGIT GPESHVLSRPEVEAEAVRAITIANQTNCLPIITKMSKSAEVAIAQARKKGTVVYG EPITASLTDGSHYKNWAKAAAFVTPSPDPPTDFLNSLLSCDGLQVTGSAHC TFNTAQKAVGKDNFTLIPGNTGTEERMSVWDKAVTGMKDNQFVATSTNAKVF NLYPRKGRIVAGDADLVIMDPDSVKTSKTHNSLEYNIPEGMCEKSGPLVTSQK KIVLEDGTLHVTEGSGRYIPRKPFDPFVKIKARSRLAEVRGPRGLYDGPVCEVSV TPKTVTASSAKTSPAKQAPPPVRLHQSGFSLSCAQIDDDNIPTRTTORIVAPPGRA NITSLG"

BASE COUNT 426 a 509 c 511 g 383 t

ORIGIN

alignment_scores:

Quality: 2339, 00 Length: 571

Ratio: 4.430 Gaps: 0

Percent Similarity: 92.469 Percent Identity: 75.482

alignment_block:

US-09-367-496-8 x HSU17279

Align seg 1/1 to: HSU17279 from: 1 to: 1829

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17

|||||

72 ATGTCTTATCAGGGGAAGAAAATATATCCACGCATCACGAGCATCGTCT 121

|||||

17 uleuileargGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34

|||||

122 TCTGATCAAGGAGGAGTAAATTTGTTAATGATCACCAGTCTGTTCTATGCAG 171

|||||

34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50

|||||

172 ACATATACATGAGATGGTTGATGATCAAGCAANTAGGAGAAATCTGATT 221

|||||


```

51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
|||||
222 GTCCAGGAGGAGTGAAGACCATCGAGGCCACTCCCGGATGGTATGCC 271
|||||
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
|||||
272 CGGAGGAATTCAGCTCCACACTCGTTCCAGATGCCCTGATCAGGAATGA 321
|||||
84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
|||||
322 CGTCTGTGATGATTTCTTCAAGGAACCAAGCGCCCTGGCTGGGGGA 371
|||||
101 ThrThrMetIleLeuAspHisValPheProAspPheGlyValSerLeuLe 117
|||||
372 ACCACTATGATCATGTGACCACTGTTCCTGAGCCTGGGACACCCCTGCT 421
|||||
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
|||||
422 CGCTGCCTTTGACCAAGTGGAGGAATGGCCGACAGCAAGTCTGCTGTG 471
|||||
134 spYrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
|||||
472 ACTACTCTCTGATGTGGACATCAGCGAGTGGCATAAGGGCATCCAGGAG 521
|||||
151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
|||||
522 GAGATGAAGCGCTGTGAGAGGATCAGCGGGTAATTCCTTCCTCGTGTA 571
|||||
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
|||||
572 CATGGCTTTCAAGATCGCTCCAGTCAACGATTCGACAGATTATGAAG 621
|||||
184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||
622 TACTGAGTGTGATCCGGGATATTGGCCCATAGCCCAAGTCCACGCAGAA 671
|||||
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLysIle 217
|||||
672 ANTGGCCACATCATTCAGAGGACGACGAGGATCTCGGATCTGGGCAT 721
|||||
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||
722 CACGGGCCCGGAGGACATGTGCTGAGCGGACCTGAGGAGCTCGAGCGCG 771
|||||
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAspCysProLeu 250
|||||
772 AAGCGGTGAATCGTGCCATCACCACCGCAACGACCAACTGCCCGCTG 821
|||||
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||
822 TATATCACCAGGTGATGAGCAAAAGCTCTGCTGAGGTTCATGCCAGGC 871
|||||
267 aLysArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
|||||
872 ACGGAAGAAGGGAATGCTGTGTATGGGAGGCCCATCATCTGCCAGCTGG 921
|||||
284 lyThrAspGlySerHisThrPheSerLysAsnTrpAlaLysAlaAla 300
|||||
922 GAACGGAGGCTCCCATTTACTGGAGCAAGAACTGGGCCAAGCTCTGCC 971
|||||
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisIle 317
|||||
972 TTTGTACCTCCACCCCTTGAGCCCTGATCCAACTCCAGACTTCT 1021
|||||
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
|||||
1022 CAACTCTCTGTCTGTGAGGACCTCCAGGTCCAGGTCACGGGCGAGTGC 1071
|||||
334 ysThrPheThrThrAlaGlnLysAlaValAlaLysAspAsnPheAlaLeu 350
|||||
1072 GCACCTTTAACATGCCAGAGGCTGTAGGAAGGACACATTCACCTG 1121
|||||
351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367

```

```

|||||
1122 ATTCGGAGGCGACCAATGGCACTGAGGAGCGGATGTCCTCATCTGGGA 1171
|||||
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
|||||
1172 CAAGGCTGTGTCTACTGGGAAGATGATGAGAACAAGTTTGTGTGTGTGA 1221
|||||
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
|||||
1222 CCGACCAATGCGACCAAGTCTTCAACCTTTACCCCGGNAAGCGCG 1271
|||||
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
|||||
1272 ATTGCTGTGGATCCGATCGGACCTGCTCATCTGGACCCCGACAGCT 1321
|||||
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1322 TAAACCATCTCTGCCAAGACACACACAGCTCTCTCGAGTACAACATCT 1371
|||||
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
|||||
1372 TTGAAGGATGTGAGTGGCGGCTCCCTGCTGCTGCTGCTGCTGCTGCTG 1421
|||||
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1422 AAGATTGCTGGAGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1471
|||||
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1472 CTACATCTCCCGGAAGCCCTTCCCTGATTTGTTTACAGCGTATCAAGG 1521
|||||
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1522 CAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1571
|||||
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1572 GGNCCGTGTGTGAAGTGTCTGTGACGCGCCAAAGACAGTCACTCCAGC 1621
|||||
517 oAaArgAlaSerCysProGlyLysIleSerValProValProValArgAsnL 534
|||||
1622 CTCGCCCAAGAGCTCTCTGCCAAGCAGGCGGCCCTGCTGCTGCTGCTG 1671
|||||
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
|||||
1672 TGCACCACTCTGATTCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1721
|||||
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1722 CCGCCGCGCACCCAGCAGGTATCTGTGCGCCCGCCCGGTGGCGTCCCAA 1771
|||||
567 nIleThrSerLeu 571
|||||
1772 CATCACCAAGCTG 1784

```

seq_name: gb_pr7:D78013

```

seq_documentation_block: 4459 bp mRNA PRI 05-FEB-1999
LOCUS D78013
DEFINITION Homo sapiens mRNA for dihydropyrimidinase related protein-2,
complete cds.
ACCESSION D78013
VERSION D78013.1 GI:1330239
KEYWORDS dihydropyrimidinase related protein-2; unc-33.
SOURCE Homo sapiens fetus brain cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4459)
AUTHORS Hamajima,N., Matsuda,K., Sakata,S., Tamaki,N., Sasaki,M. and
Nonaka,M.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1995) to the DDBJ/EMBL/GenBank databases. Naoki

```



```

417 rlyslleileserAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1525 TAAACCATCTCTGCCAAGACACACAGCTCTCTCGAGTACAACATCT 1574
434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1575 TTGAAGGCATGAGTGGCGCGCTCCCACTGGTGGTTCATCAGCCAGGG 1624
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 467
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1625 AGATGTCTCTGGAGACCGCCCTCGATGTCACCGAAGGCTCTGGAGC 1674
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1675 CTACATTCCTCCGGAAGCCCTTCCTGATTTGTTACAAAGCGTATCAAG 1724
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1725 CAAGGAGCAGGCTGCTGAGCTGAGAGGGTCTCTGCGCCCTGATGAC 1774
501 GlyProValHisGluValMetProAlaLysProGlySerGlyAlaPr 517
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1775 GGACCTGTGTGAAGTGTCTGTGCGCCCAAGACAGTCACTCCAGCCTC 1824
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1825 CTCGGCCACAGCTCTCTGCCAAGCAGCAGGCCCCACCTGTCCGGAACC 1874
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1875 TGCACCACTGTGATTCAGTTGCTGTGGTCTCAGATTGATGACACATT 1924
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1925 CCGCGCGCACCCACCGCTGCTGTGGCGCCCGCGGTGGCGGTGCCAA 1974
567 nileThrSerLeu 571
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1975 CATCACCAAGCTG 1987
seq_name: gb_pr10:HSU97105

seq_documentation_block:
LOCUS HSU97105 5421 bp mRNA PRI 17-MAR-1998
DEFINITION Homo sapiens N2A3 mRNA, complete cds.
ACCESSION U97105
VERSION U97105.1 GI:2967518
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5421)
AUTHORS Gu,J.R., Chen,Y. and Zhou,J.
TITLE A cDNA clone highly expressed in human brain and deleted in liver
cancer
JOURNAL Unpublished
2 (bases 1 to 5421)
AUTHORS Gu,J.R., Chen,Y. and Zhou,J.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1997) National Laboratory for Oncogenes and
Related Genes, Shanghai Cancer Institute, 25/Ln 2200, Xie-Tu Road,
Shanghai, Shanghai 200032, P. R. China
3 (bases 1 to 5421)
AUTHORS Gu,J.R., Chen,Y. and Zhou,J.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1998) National Laboratory for Oncogenes and
Related Genes, Shanghai Cancer Institute, 25/Ln 2200, Xie-Tu Road,
Shanghai, Shanghai 200032, P. R. China
REMARK Nucleotide and amino acid sequence updated by submitter
COMMENT On Mar 17, 1998 this sequence version replaced gi:2109272.
FEATURES

```

SOURCE

```

1. 5421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
/dev_stage="fetus"
1337.3055
/codon_start=1
/product="N2A3"
/protein_id="AAC05793.1"
/db_xref="GI:2967519"
/translation="MSYQKKNIPIRISDRLLIKGGKIYNDQSFYADLYMEDGLIKO
IGENLVPGGVKATIEAHSRMVTPGGIDVHTRFQMDQGMTSADDFQGTAKAALAGTT
MIIDHVPFGTSLAFAFDQWREWADSKSCDYSLHVDISEWHKGIQEEAEALVKDHG
VNSFLVMAFKDRFOULTDQIYEVLSVIRDIQAIQVHAENGDIIEAQORILDLGIT
GPEGHVLSRPEEVEAEAVNRAITIANOTNPLYITKVMSSKSAEVLQAQKRGTVVYG
EPITASLGDGSHYKSNKAKAAFTVSPPLSPDPPTDFLNSLLSCGLQVGTSAHC
TFNTAKRAGKDNFTLIPGTNGTTERMSVINDKAVVTKMDENOFVAVTSTNAAKVF
NLYPRKRGIAVGSDALVWDPSVATISAKTHNSLSLEINIEFGMECRSPVLVVISQG
KIVLEDTLHVTEGSGRYTPRPFPPDFYKRIKARSLRAELRGVPRGLIDGPPCEVSY
TPKTVTPASSAKTSPAKQOAPPVVRNLHQSGFSLSGAOIDDINPRTRTQRIVAPPGGA
NITSLG"

```

CDS

```

BASE COUNT 1390 a 1269 c 1209 g 1553 t
ORIGIN

```

```

alignment_scores:
Quality: 2339.00 Length: 571
Ratio: 4.430 Gaps: 0
Percent Similarity: 92.469 Percent Identity: 75.482

alignment_block:
US-09-367-496-8 x HSU97105 ..
Align seg 1/1 to: HSU97105 from: 1 to: 5421

1 MetSerPheGlnGlyLysSerIleProArgIleThrSerAspArgLe 17
1337 ATGCTCTTATCAGGGGAAGAAATAATTCACGCATCAGGAGGATCGTCT 1386
17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
1387 TCAGATCAAGAGGAGTAAATTTGTTAATGATGACCAAGTCTTCTATGAC 1436
34 spValHisValGluAspGlyLeuIleGlyGlnIleGlyGluAsnLeuIle 50
1437 ACATATACATGGAAGATGGGTTTATCAAGCAATAGGAGAAATCTGAT 1486
51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
1487 GTGCCAGGAGGAGTGAAGACCATCGAGGCCCACTCCCGGATGCTGATCCC 1536
67 OGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
1537 CGGAGGAATTCAGCTGCACACTCGTTTCCAGATGCTGTGATCAGGGAATGA 1586
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
1587 CGTCTGCTGATGATTTCTTCCAAAGGAACCAAGCGGCCCTGGCTGGGGGA 1636
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
1637 ACCACTATGATCATTTGACCATGTTTCTTCCAGCTGGGACAAAGCCTGCT 1686
117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
1687 CGCTGCCTTTGACCAGTGGAGGAAATGGGCGCACGACGAAGTCTCTGCTGTG 1736
134 spYrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
1737 ACTACTCTCTGATGTGGACATCAGCGAGTGGCATAAGGGATCCAGAG 1786
151 GluLeuGluAlaLeuValLysGlyValLysGlyValAsnSerPheLeuValPh 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

1787 GAGATGGAGCGCTTGTGAAGATCATCGGGTAATTCCTTCCTCGTGTA 1836
167 eMetalaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
1837 CATGGCTTCAAGATCCCTCCAGCTAACGATTCAGATTTATGAAG 1886
184 lePheSerIleleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
1887 TACTGAGTGTATCGGGATATTCGGCCATAGCCCAAGTCCAGCCAGAA 1936
201 AsnGlyAspIleValGluGluGlnGlnLysArgLeuLeuGluLeuGlyI 217
1937 AATGGCGACATCATTCAGAGGACACAGAGGATCCTGGATTCGGGCAT 1986
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
1987 CACGGGCGCGAGGACATGTGTGAGCCGACCTGAGGAGGTCTGAGGCCG 2036
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
2037 AGCGGTGAATCGTGGCATCACCATCGCCCAACAGACCAACTGCCGCTG 2086
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
2087 TATATCACCAAGGTGATGACAAAGCTGTCTGAGTCACTGCCCGAGGC 2136
267 aLysArgArgGlyValValValpHeGlyGluProIleThrAlaSerLeuG 284
2137 ACGGAAGAGGAACTGTGTGTATGGCGAGCCCATCATCGCAGCTTG 2186
284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaA 300
2187 GAACGGAGGCTCCCATTAATCTGAGGACAGAACTGGCCAGGCTGTGCC 2236
301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
2237 TTGTCTACCTCCCGCCCTGTGAGCCCTGTATCCAACTCCAGACTTCT 2286
317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
2287 CAACCTCTGTCTCTCTGGAGACCTCCAGGTCACGGCAGGTGCCAT 2336
334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
2337 GCAGCTTTAACTACTGCCAAGAGGTGTAGGAAAGGACAACTTACCCCTG 2386
351 IleProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpG1 367
2387 ATTCGGAGGCGACCAATGGCACTGAGGAGCGGATGTCGTCTCTGGGA 2436
367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
2437 CAAGGCTGTGGTCACTGGGAAGATGGATGAGAACCACTTGTGGTGTA 2486
384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
2487 CCAGCCCAATACACCAAGTCTTCAAGCTTTACCCCGGAAAGCGCG 2536
401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
2537 ATTGCTGTGGATCGATCGGACCTGTCTATCTGGGACCCCGCAGCGGT 2586
417 lyllelleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
2587 TAAACCATCTCTGCAAGACACACACAGCTCTCTCGAGTACACATCT 2636
434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
2637 TGAAGCATGAGTGGCGCGGCTCCCGCTGGTGTCTATCAGCCAGGGG 2686
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
2687 AAGATTGTCTGGAGGAGGCGCCCTGCATCTCACCGAAGGCTCTGACG 2736

467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
2737 CTACATCCCGGAAGCCCTTCCCTGATTTGTTTACAAAGCGTATCAAG 2786
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
2787 CAAGGACGAGCTGGCTGAGTGTGAGAGGGTTCCTCGTGGCTGTATGAC 2836
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
2837 GGACCTGTGTGAAGTGTCTGTGACGCCCAAGACAGTCACTCCAGCCTC 2886
517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
2887 CTCGGCCCAAGACGCTCTCTGCCAAGCAGCAGGCCCTCTCGGAAAC 2936
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
2937 TGCACCACTGTGGATTCAGTTGTCTGTGTCTGATGATGACACAT 2986
551 AlaArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
2987 CCCGCGCGCACCCACCCAGCGTATCGTGGCGCCCGCGTGGCGTGCAC 3036
567 nileThrSerLeu 571
3037 CATCACCGAGCTG 3049
seq_name: gb_patl:A87370
seq_documentation_block:
LOCUS A87370 1817 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9837192.
ACCESSION A87370
VERSION A87370.1 GI:6736126
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1817)
AUTHORS Byk,T. and Belin,M.
TITLE USE OF ULIP PROTEINS IN THE DIAGNOSIS AND THERAPY OF CANCER AND
PARANEOPLASTIC NEUROLOGICAL SYNDROMES
JOURNAL Patent: WO 9837192-A 1 27-AUG-1998;
BYK TAMARA (CH); BELIN MARIE FRANCOISE (FR)
FEATURES
Location/Qualifiers
source 1..1817
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 426 a 506 c 511 g 374 t
ORIGIN

alignment_scores:
Quality: 2325.00 Length: 571
Ratio: 4.429 Gaps: 0
Percent Similarity: 91.944 Percent Identity: 75.306
alignment_block:
US-09-367-496-8 x A87370 ..
Align seg 1/1 to: A87370 from: 1 to: 1817
1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgIle 17
23 ATGCTTATCAGGGGAGAGAAATATTCACCCATCAGCAGGATCGTCT 72
17 uleuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaA 34
73 TCTGATCAAGGTGGCAAGATGTGATGATGATGATGATGATGATGATGAT 122

34 spValHisValGluAspGlyLeuIleGlnLysGlnIleGlyGluAsnLeuIle 50
 123 ACATATACATGGAAGATGGTTGATCAAGCAATAGGAGAAACCTGATT 172
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 173 GTACCAGGAGGGTGAAGACCATCAAGCCACTCCAGAAATGGTATTCC 222
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 223 CGGAGGAATTGACGTGCTACTCTCCAGATGCTGACCAAGGAATGA 272
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 273 CATCCGCTGATGACTTCTCCAGGGAACCAAGCGGCGCTGCGCGGGA 322
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 323 ACCACCATGATCATGACCATGTTCTTCTGAGCCCGGACGAGCCTATT 372
 117 uAlaAlaTyrGluGlnTyrArgGluAlaAspSerAlaAlaCysCysA 134
 373 GCGTCCCTTTGATCACTGAGGAGTGGCTGACAGCAAGTCTCTGCTGG 422
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 423 ACTATTCTGCTGCACGTGGACATCACTGAGTGGCAAGGGCATCCAGGAG 472
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 473 GAGATGGAAGTCTGCTGGAAGGACCAAGCGGGTAAACTCTTCTCGTGT 522
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
 523 CATGCTTTCAAGATGATTCACACTGACGGATTCACGATCTATGAGAG 572
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 573 TGCTGAGCGTATCCGGATATCGGTGCTGATAGTCAAGTCCACGACGAG 622
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
 623 AATGGTGACATCATTTGCTGAGGACACAGCAGAGGATCTTGGATCTGG 672
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 673 CACGGGCCCCGAGGACACAGTGTGGCCGCGCAGAGGAGTCTGAGGCTG 722
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 723 AAGCTGTCAACCGTCCATCACTATTGCGCAACACGACCAACTGCCCTCTG 772
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 773 TATGTCACCAAGTGTATGCCAAGAGTGGCGTGAAGTCATGCTCAGGC 822
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
 823 ACGGAAGAGGGAAGTGTGGTGTATGGTATGAGCCCATCAGCGCCAGCTGG 872
 284 lYThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 873 GGACTGATGGCTCTCATTTACTGGAGCAAGAACTGGGCAAGGCTCGCGCC 922
 301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
 923 TTGTGACCTCCACCTTGGAGCCCGCCACCACTCCAGACTTTCT 972
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 973 CAACTCGTGTCTGCTGTGAGACCTCCAAAGTCACTGCGAGTGGCCACT 1022
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350

1023 GCACCTTCAACACTGCCAGAGGCTGTGGGAAGCAACACTTCACCTTG 1072
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGl 367
 1073 AUTCCGAGGGCACCACCAACGCACTGAGGAGCGGATGCTCTCATTTGGGA 1122
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValT 384
 1123 TAAAGCTGTGGTCACTGGGAAGATGATGAGAATCAGTTTGTGGCTGTGA 1172
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1173 CCAGCACCAACGACGAGCAAAAGTCTTCAACCTTTACCCCGGAAAGCTCG 1222
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1223 ATCTCGTGGGATCTGATGCTGACTTGGTCTATCTGGGACCTGCACAGTG 1272
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1273 GAAGACCATCTCTGCCAAGACACACACAGTGTCTTTGAGTACAACATCT 1322
 434 heGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGly 450
 1323 TTGAAGGCATGGAGTCTCGCGGCTCCCACTGGTGTCTATCAGCAGGCG 1372
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1373 AAGATTGTCTCGAGGACGCACTTCATGCTCACTCAAGGCTCAGGACG 1422
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
 1423 CTACATTTCCCGGAGCCCTTCCCTGACTTTGTGTACAAACGATCAAG 1472
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1473 CAAGGAGCAGGCTGGCTGAGCTGAGAGGGTCCCTCGTGGCCCTGTATGAC 1522
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1523 GGACCGGTATGCGAGGTGTCTGTGACGCCCAAGACGGTGTACTCCAGCCTC 1572
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1573 ATCAGCTAAGACATCCCTGCCAAGCAGGACGACCCCTGTTCGGAACC 1622
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
 1623 TGCACCACTCTGGATTACGCTTGTCTGTGCTCAGATTGACGACACATTT 1672
 551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
 1673 CCCTGGCGCACCAACGAGGATCTGTGGCAGCCCTTGGTGGCGGTGCCAA 1722
 567 nIleThrSerLeu 571
 1723 CATCACCAAGCTTG 1735

...

Tue Jul 31 13:08:21 2001

OM of: US-09-367-496-8 to: EST:* out_format : pfs

Date: Jul 30, 2001 12:07 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame-p2n.model -DEV=xlip
-O=/cgn2.1/USPTO.spool/US09367496/runat_30072001_081220_27362/app_query.fasta_1.637
-DB=EST -FMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPOP=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09367496 -CGN1_1_3359 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-367-496-8

Query length: 572

Database: EST*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 1362.640000

score_list:

Sequence	strd	Orig	zScore	Escore	Len	Documentation
gb_hlc:AK004822	+	1518.50	2772.79	2.5e-145	2109	AK004822 Mus musculus adult male
gb_hlc:AK004899	+	1515.00	2764.41	7.3e-145	2471	AK004899 Mus musculus adult male
gb_est29:AL535253	+	1218.00	2226.32	6.9e-115	989	AL535253 AL535253 LTI_FL013.FBI
gb_est29:AL535895	+	1103.00	2014.50	4.3e-103	925	AL535895 AL535895 LTI_FL013.FBI
gb_est29:AL535817	+	1101.50	2011.86	6.1e-103	925	BF530817 602072175F1 NCI-CGAP-FBI
gb_est29:AL535819	+	1091.00	1992.38	7.4e-102	920	AL535819 AL535819 LTI_FL013.FBI
gb_est29:AL535854	+	1068.00	1951.88	1.3e-99	778	AL535854 AL535854 LTI_FL013.FBI
gb_est11:AA778518	+	1046.00	1913.13	1.9e-97	663	AA778518 af85905.sl Soares_test
gb_est22:AL569716	+	1029.00	1882.66	9.5e-96	612	AL569716 tn05a06.xl NCI-CGAP-FBI
gb_est51:AW717374	+	1025.00	1875.67	2.3e-95	592	AW717374 hn55g08.xl NCI-CGAP-FBI
gb_est29:AL535160	+	1035.00	1868.77	5.7e-95	1049	AL535160 AL535160 LTI_FL013.FBI
gb_est29:AL536357	+	1015.00	1849.53	6.7e-94	1117	AL536357 602390746F1 NIH_MGC_94
gb_est38:BG293352	+	995.00	1818.32	3.7e-92	697	BG293352 602390746F1 NIH_MGC_94
gb_est30:AU118714	+	992.50	1812.36	7.9e-92	773	AU118714 AU118714 HEMBAL Homo s
gb_est29:AL537474	+	985.00	1798.32	4.8e-91	784	AU537474 AU537474 LTI_FL013.FBI
gb_est42:AW090187	+	984.50	1799.19	4.3e-91	676	AW090187 xc92b10.xl NCI-CGAP-FBI
gb_est29:AL536884	+	980.00	1786.50	2.2e-90	971	AL536884 AL536884 LTI_FL013.FBI
gb_est29:AL536824	+	959.00	1752.27	1.7e-88	663	AL536824 AL536824 BP Chicken Emb
gb_est27:AJ394827	+	957.00	1748.45	2.9e-88	670	AJ394827 AJ394827 dkf42426 Gallu
gb_est30:AU124654	+	946.00	1727.41	4.2e-87	710	AU124654 AU124654 NT2RM4 Homo s
gb_est04:BF206042	+	930.50	1695.63	2.5e-85	920	BF206042 601868675F1 NIH_MGC_17
gb_est29:AL536988	+	928.00	1691.48	4.3e-85	885	AL536988 AL536988 LTI_FL013.FBI
gb_est79:BF793111	+	925.50	1686.57	8.0e-85	906	BF793111 601591619F1 NIH_MGC_7
gb_est30:AU080879	+	924.50	1685.85	8.8e-85	825	AU080879 AU080879 Sugano mouse
gb_est30:AU126219	+	924.00	1686.24	8.3e-85	740	AU126219 AU126219 NT2RP1 Homo s
gb_est30:AW950945	+	920.00	1679.99	1.9e-84	673	AW950945 EST363015 MAGE resequ
gb_est104:D88457	+	910.00	1650.49	8.2e-83	1678	D88457 D88457 Mouse ICR Testis
gb_est30:AU121198	+	898.50	1638.20	4.0e-82	797	AU121198 AU121198 HEMBAL Homo s
gb_est98:BG295351	+	893.50	1627.54	1.5e-81	896	BG295351 602390746F1 NIH_MGC_94
gb_est43:AL161035	+	890.00	1623.26	2.7e-81	747	AL161035 AU78404.yl Schneider f
gb_est29:AL538476	+	885.00	1611.91	1.2e-80	890	AL538476 AL538476 LTI_FL013.FBI
gb_est29:AL538144	+	881.00	1604.03	3.2e-80	926	AL538144 AL538144 LTI_FL013.FBI
gb_est29:AL535744	+	879.00	1598.04	4.6e-80	870	AL535744 AL535744 LTI_FL013.FBI
gb_est23:AL1663726	+	878.00	1590.49	6.8e-80	961	AL1663726 ud47b12.yl Soares_mamm
gb_est29:AL535814	+	876.50	1593.33	9.7e-80	956	AL535814 AL535814 LTI_FL013.FBI
gb_est29:AL535812	+	876.00	1591.80	5.5e-79	682	AL535812 AL535812 LTI_FL013.FBI
gb_est101:BG471608	+	867.00	1581.82	4.6e-79	860	BG471608 602513179F1 NIH_MGC_10
gb_est29:AL585536	+	866.00	1577.47	9.5e-79	840	AL585536 AL585536 BP Chicken Emb
gb_est29:AL525393	+	866.00	1567.06	3.6e-78	857	AL525393 AL525393 LTI_FL013.FBI
gb_est31:AU132127	+	852.00	1549.55	3.4e-77	995	AU132127 AU132127 NT2RP3 Homo s
gb_est29:AL534424	+	848.00	1547.41	4.5e-77	943	AL534424 AL534424 LTI_FL013.FBI
gb_est29:AL586393	+	848.00	1547.41	4.5e-77	943	AL586393 AL586393 BP Chicken Emb

gb_est91:BF685809 + 846.50 1541.48 9.6e-77 836 + BF685809 602140492F1 NIH_MGC
gb_est29:AL528054 + 846.00 1539.46 1.2e-76 915 + AL528054 AL528054 LTI_FL013
gb_est81:BE952303 + 844.00 1538.34 1.4e-76 739 + BE952303 UI-M-CD0-ayl-b-04-0
gb_est29:AL535518 + 844.00 1535.98 2.0e-76 899 + AL535518 AL535518 LTI_FL013

seq_name: gb_hlc:AK004822

seq_documentation_block: 2109 bp mRNA HTC 08-FEB-2001
LOCUS AK004822
Mus musculus adult male lung cDNA, RIKEN full-length enriched
library, clone:1200017110, full insert sequence.

ACCESSION AK004822.1 GI:12836294

VERSION CAP trapper.

Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA,

clone:1200017110.

Mus musculus

Mus musculus

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Methods Enzymol. 303, 19-44 (1999)

2 (sites)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

3 (sites)

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,

Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T.,

Harada, A., Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T.,

Kashiwagi, K., Fujii, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsui, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system -384-format

sequencing pipeline with 384 multipipette sequencer

Genome Res. 10 (11), 1757-1771 (2000)

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 2109)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,

Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,

Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,

Kojima, Y., Konno, H., Kouda, S., Koyama, K., Kurihara, C., Matsuyama, T.,

Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,

Okido, T., Owa, C., Saito, H., Saito, R., Sakai, K., Sano, H.,

Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A.,

Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

Fax: 81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

COMMENT

/translation="MAPQRLIRGRIVNDDFSQVADLVEDGVVRLGRDLLPPED
ASRGRIILDAAGKLVLPGLIDHTHMQPFPMGSSQVDDFYQCTKAALAGGTTMTIDFA
IPKSGSLIEAFETWRNADPKVCCDYSILHVAVTWMSDKVKEEMKTLARDGVNSFKM
FMAYKGLYVDOEOLYAAFSOCKEIGAIAOVHAENDLIIAEGAKTLALGITGPBGHE
LCRPVAVAEATLRAITASAVNCPPLYVHVMSKAARVADARRAGNVVYGEPTAAG
LGTGROVYSEWSESHAAHVMPPLRPDLPLPGELMDLLANGDLTTTGGSDNCTFNQC
KALGKDDTDKIPNGVNGVEDRMSVLEWEGVHSGKMDENRFVAVTSTNAKIFNLYPKK
GRIAVGDDKIVINDPEATRISAKTHHQAVNFNIFEGMVCHGVPLVTTISGRVVEA
GVFNVTAGHGKFIPIRQPAEYIYKRIRKORDQCTPVPVKRAPHYKGEVTLKARETKED
DTAGTRMQGHS"
2448..2453
/note="putative"
2471
polyA_signal
polyA_site
BASE COUNT 682 a 583 c 636 g 570 t
ORIGIN

alignment_scores:
Quality: 1515.00 Length: 492
Ratio: 3.788 Gaps: 1
Percent Similarity: 81.301 Percent Identity: 58.537

alignment_block:

US-09-367-496-8 x AK004899 ..

Align seg 1/1 to: AK004899 from: 1 to: 2471

16 ArgLeuLeuLeuArgGlyGlyValArgIleValAsnAspGlnSerPhe 32
|||||
124 GCACCTTCATCCCGGGGGTGCATGTCATGACGACTTCTCACAGGT 173
32 rAlaAspValHisValGluAspGlyLeuIleValGlnIleGlyGluAsnL 49
|||||
174 GCCGATGCTAGTGGAGGACGGCGTGGCGCGCTGGAGGGGACC 223
49 euIleValPro.....GlyGlyIleHisThrIleAlaHis 61
|||||
224 TGCTGCTCCTCAGGACGACATCTCGGGGCTGCGAATCTTAGATGCACGG 273
62 GlyLeuMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMe 78
|||||
274 GGCAAGCTGCTCGCCCGGGGGCATCGACACGACGACATGCAAT 323
78 tProValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysA 95
|||||
324 CCCGTTTGGGCTCGCAGTCAGTGGAGGACTTCTACAGGGCACCAGG 373
95 laAlaLeuAlaGlyGlyThrMetIleLeuAspHisValPheProAsp 111
|||||
374 CTGCTTTGGCAGGAGGACCAACCATGATGATTTGTCATTTCTCTCAG 423
112 ThrGlyValSerLeuLeuAlaAlaValArgGlnThrArgGluArgAla 128
|||||
424 AAAGGCAAGCTCCCTCATCGAGGCTTTGACACTTGGCCAACTGGGCAGA 473
128 pSerAlaAlaCysCysAspTyrSerLeuHisValAlaSpilleThrArgtrp 145
|||||
474 CCGCAAGTCTGCTGACTATACCTGTCATGTCGTCGTCGTCGTCGTCG 523
145 isGluSerIleLysGluLeuGluAlaLeuValLysGluLysGlyVal 161
|||||
524 GCGACAAGGTAAAGAAAGAAATGAAACCTTGCCCGAGATAAAGCGGT 573
162 AsnSerPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAs 178
|||||
574 AACTCTTTCAAGATGTTTATGGCTACAAAGGCTCTGTACATGTGTGCA 623
178 pSerGlnMetTyrGluIlePheSerIleIleArgAspLeuGlyAlaLeuA 195
|||||
624 CGAGCAGCTGTATGGCCCTTCTCAGTGGCAAGGAGATAGAGGATG 673
195 laGlnValHisAlaGluAsnGlyAspIleValGluGluGlnLysArg 211

|||||
674 CTCAGGTGCAGCGCGAGATGGAGATTGATTCAGAGGGGAGCAAGAAG 723
212 LeuLeuGluLeuGlyIleThrGlyProGluGlyHisValLeuSerHisPr 228
|||||
724 ACCTGGCAGTGGGATTAACAGGCCCTTGAGGGGCACAGCTGTGCCGTCC 773
228 oGluGluValGluAlaGluAlaValTyrArgAlaValThrIleAlaLysG 245
|||||
774 AGAAGCAGTGGAGGACAGGCCACCTGAGAGCCATCACCATTAGCCAGCG 823
245 lnAlaAsnCysProLeuTyrValThrLysValMetSerLysGlyAlaAla 261
|||||
824 CTGTGAAGTGCCTTTATACGTGCTGATGATGAGCAAAATCTGCAGCG 873
262 AspAlaIleAlaGlnAlaLysArgArgGlyValValValPheGlyGluPr 278
|||||
874 AAGTGTGTAGCGATGCGAGGAGAGAGCAAAATGCTGTATGGAGACC 923
278 orIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrpSerLysAsnT 295
|||||
924 AATTCAGCGGCTCGGCACCTGATGGCAGACAGTACTGGAGTGAAGAAT 973
295 rpAlaLysAlaAlaAlaPheValThrSerProProValAsnProAspPro 311
|||||
974 GGAGCCATGCGCCACCATGTCATGGGTCCCTCCCTGAGACCCGATCCT 1023
312 ThrThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnVa 328
|||||
1024 TTAACACCTGGCTTTCTTATGATCTGTGGCTAATGGGATTTCACCA 1073
328 lThrGlySerAlaHisCysThrPheThrThrAlaGlnLysAlaValGlyL 345
|||||
1074 AACAGGAGTGCACACTTTTAAACCTTGCACAAAGCTCTAGGGA 1123
345 ysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGluGluArg 361
|||||
1124 AGGATGACTTCAATGAATGCCAATGGGTGAATGGTTCGAGGACAGG 1173
362 MetSerMetValTyrGluLysCysValAlaSerGlyLysMetAspGluAs 378
|||||
1174 ATGTCGGTGATATGGGAAAGGGCGTCACAGTGTGTAATAATGGATGAAA 1223
378 nGluPheValAlaValThrSerThrAsnAlaAlaLysIlePheAsnPheT 395
|||||
1224 CAGTTTGTGGCAGTTACCACACAAACGACGCAAAATCTTTAATCTTT 1273
395 yrProArgLysGlyArgValAlaValGlySerAspAlaAspLeuValIle 411
|||||
1274 ATCCGAAAAGGAAGAAATAGCTGTAGGCTCAGATGCTGACATTTGTATC 1323
412 TrpAsnProLysAlaThrLysIleIleSerAlaLysThrHisAsnLeuAs 428
|||||
1324 TGGACCCAGAACCAACAGGAGATCTCAGCCAAACTCATCATCAGGC 1373
428 nValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaProAlaV 445
|||||
1374 CGTTAATCTCAACATCTTCGAGGGCATGTTTGTACGGGTGCCCTGG 1423
445 alValIleSerGlnGlyArgValAlaLeuGluAspGlyLysMetPheVal 461
|||||
1424 TGACTATTTCAAGAGCAGAGTGGTGTATGAGCAGGTTGTTTCAACGTC 1473
462 ThrProGlyAlaGlyArgPheValProArgLysThrPheProAspPheVa 478
|||||
1474 ACAGCAGGACATGGGAAGTTTATTCGCCCAACCTTTGCCCAATACAT 1523
478 lTyrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValP 495
|||||
1524 TTACAAACCAATCAACAGCAGGACGACACCTGTCACACCTGTGCCGTGA 1573
495 roArgGlyLeuTyrAspGlyProVal 503
|||||

```

1574 AGCGGGCCCTTACAGGAGATC 1599
seq_name: gb_est29:AL535253

seq_documentation_block:      989 bp    mRNA          EST       13-FEB-2001
AL535253 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF008YF05 5
LOCUS DEFINITION prime, mRNA sequence.
AL535253
AL535253 AL535253.1 GI:12798746
ACCESSION EST.
VERSION human.
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 989)
Li.W.B., Gruber.C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..989
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF008YF05"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 224 a 276 c 296 g 192 t
ORIGIN

alignment_scores:
Quality: 1218.00 Length: 315
Ratio: 4.289 Gaps: 0
Percent similarity: 90.159 Percent identity: 69.841

alignment_block:
US-09-367-496-8 x AL535253 ..
Align seg 1/1 to: AL535253 from: 1 to: 989
60 AlahisgLyLeuMetValLeuProGlyGlyValAspValHisThrArgLe 76
3 GCCAACGGCGGTATTTCGGAGGTATTGCATCAACGTTACT 52
76 uclMetProValLeuGlyMetThrProAlaAspPheCysGlnGlyt 93
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
53 GCAGAAGCCCTCCAGGGGATGACTGCCGCTGATGACTCTTCCAAGGGA 102
93 hrLySaLaLeuAlaGlyThrThrMetLeuAspHisValPhe 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 CCAGGGCGCACCTGGTGGGGGACCACGATGATCATGACCATGTTT 152
110 ProAspThrGlyValSerLeuLeuAlaAlaAtyrGluInTrpArgGluAr 126
||||| ||| ||| ||||| ||||| ||||| ||||| |||||
:53 CTGAACCTGGTCCAGCTACTGACCTCTTTCAGAAAGTGCCACGAAGC 202
126 gAlaAspSerAlaAlaCysCysAspTyrSerLeuHisValAspLeThrA 143

```


Tue Jul 31 13:08:21 2001

/tissue_type="anaplastic oligodendroglioma with lp/19q loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
197 a 271 c 311 g 136 t

BASE COUNT
ORIGIN

alignment_scores:
Quality: 1101.50 Length: 269
Ratio: 4.337 Gaps: 4
Percent Similarity: 94.424 Percent Identity: 91.822

alignment_block:
US-09-367-496-8 x BF530817 ..

Align seg 1/1 to: BF530817 from: 1 to: 915

50 lilevalProGlyGlyLeuHisThrIleAspAlaHisGlyLeuMetValLe 66
|||||
3 ATCGTCCCTGGGGCATCAAGACCATGACGCCACGGCTG.ATGCTCT 51
66 uProGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyM 83
|||||
52 TCCTGGTGGGTTGAGCTCCACAAAGGCTGCAGATGCTCTCTCGGGCA 101
83 etThrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGly 99
|||||
102 TGACACCGGCTGACGACTCTGTGAGGCGACCAAGCGAGCGCTAGCAGGA 151
100 GlyThrThrMetLeuAspHisValPheProAspThrGlyValSerLe 116
|||||
152 GGAACCAACCATGATCTGGACACAGCTCTTCCCGACACGGGTGTGAGCT 201
116 uLeuAlaAlaTyrgluGlnTrpArgGluArgAlaAspSerAlaAlaCysC 133
|||||
202 GCTGGCGGCTTACGACGAGTGGGGGAGCGGGGAGCGGCGGCGGCGGCT 250
133 ysAspTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLys 149
|||||
251 GCGACTACTCTCCGCTGACATCATCCCGATGCGCATGAGCATCAAG 300
150 GluGluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuVa 166
|||||
301 GAGGAGCTGGAGGCTCTGCTCAAGAGAGAGGGTGTGAACCTCTCTCTGGT 350
166 lPheMetAlaTyrlLysAspArgCysGlnCysSerAspSerGlnMetTyrG 183
|||||
351 CTTTCATGTCATCAAGGACCGGTGTCAGTGCAGCGACAGCGAGTACG 400
183 luIlePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAla 199
|||||
401 AGATCTTCAGCATCATCCGGGACCTTGGGGCTTGGCCCGAGGTGACGCT 450
200 GluAsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlu 216
|||||
451 GAGACGGGGGACATCTGGAGGAGGAGCAGAAAGCGGTGCTGGAGCTCGG 500
216 ylleThrGlyProGluGlyHisValLeuSerHisProGluGluValGluA 233
|||||
501 CATCACTGGCCCGGAGGCGACGTGCTCAGCCACCCCGAGGAGGTGGAGG 550
233 laGluAlaValTyrgAlaValThrIleAlaLysGlnAlaAsnCysPro 249
|||||
551 CTGAGGGGGGTGTACCGAGCTGTCCACCATCGCCCAAGCAGCGC.AACTGCCCG 599
250 LeuTyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaG 266
|||||
600 CTGTACGTCACCAAGGTGATGACCAAGGGGGCGGC.GACGCCATCGCTCA 648

266 nAlaLysArgArgGlyValVal ValPheGlyGluProIleThrAlaSer 282
|||||
649 GCCAAGCGCAGAGGGTGTGCGGTA...GGGAGGCCCATCACCGC.AGC 694
283 LeuGlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaA 299
|||||
695 CTGGGCGACCGAGGTTTACACACTTCTGAGCAGACTGG...CCAGGCGC 741
299 aAlaPheValThrSerProValAsnProAspProThrThrAlaAspH 316
|||||
742 CAGCCTTCGTACATCACCCCTGGTA...CCAGACCAACAAGGGGACACT 788

316 lsLeu 317
|||
789 CACTG 793

seq_name: gb_est28:AL518319

seq_documentation_block:
LOCUS AL518319 920 bp mRNA EST 13-FEB-2001
DEFINITION AL518319 LTI_NFL011_NBC1 Homo sapiens cDNA clone CSODA009YA18 5
prime, mRNA sequence.

ACCESSION AL518319
VERSION AL518319.1 GI:12781812
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 920)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

FEATURES

source

1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODA009YA18"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"

BASE COUNT 220 a 276 c 256 g 166 t 2 others
ORIGIN

alignment_scores:
Quality: 1091.00 Length: 305
Ratio: 4.041 Gaps: 0
Percent Similarity: 88.525 Percent Identity: 66.885

alignment_block:

US-09-367-496-8 x AL518319 ..

Align seg 1/1 to: AL518319 from: 1 to: 920

258 LysGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValva 274
|||||

```

3  AAGAGTGCAGCCGACATCATCTGCTGGCCAGGAAAGGCGCCCTAGT 52
274  lPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrT 291
53  TTTTGGAGAGCCCATTTGCCGCGAGCTGGGACCGATGCCACCATTTACT 102
291  rpSerLysAsnTrpAlaLysAlaAlaPheValThrSerProVal 307
103  GGAGCAAGAACTGGCCAAAGCTGGGGTTCGTGACTTCCCTCCCTG 152
308  AsnProAspProThrThrAlaAspHisLeuThrCysLeuLeuSerSerG 324
153  AGCCCGGAGCCCTACACAGCCGCTACTTACTGACCTCCCTACTGGCTTGG 202
324  yAspLeuGlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnL 341
203  GGACTTGCAGTGCAGCAGCGGCGCTACTTCCCTCAGCAGCTGCCGAGA 252
341  ysAlaValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGly 357
253  AGCGGCTGGGCAAGGACACTTTACCTGTATCCCGAGGGGTGTCAACGGG 302
358  IleGluGluArgMetSerMetValTrpGluLysCysValAlaSerGlyLy 374
303  ATAGAGGAGCGGATGACGGTCTGCTGGACAGCGGTGGCTACTGGCAA 352
374  sMetAspGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysI 391
353  AATGGATGAGAACCACTTGTGCTGTCCACGACCAATGACGACCAAGA 402
391  lePheAsnPheThrProArgLysGlyArgValAlaValGlySerAspAla 407
403  TCTTTAACTGTACCAAGAAAGGCGGATTCGCTGGCTCGGATGCC 452
408  AspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysTh 424
453  GACGTGCTCATCTGGACCCGACCAAGTTGAAGACCATAACAGCCAAAG 502
424  rHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArg 441
503  TCACAAGTGGCGGTGGAGTACAACTTCGAGGGTATGAGTGCCACG 552
441  lyAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGly 457
553  GCTCCCCACTAGTGTATCATCAGCCAGGCAAGATCGCTTTGAAGACGA 602
458  LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPh 474
603  AACATCAACGTCAACAAGGCGATGGCGCTTCATTCCGCGGAAGCGTT 652
474  eProAspPheValThrLysArgIleLysAlaArgAsnArgLeuAlaGluI 491
653  CCGGAGCACCTGTACGAGCGGCTCAAAATCAGGAATAAGTTTGGAT 702
491  leHisGlyValProArgGlyLeuTyrAspGlyProValHisGluValMet 507
703  TGCAAGGGTTCACAGGGCATGTATGACGGTCTGTGTACAGGATACCA 752
508  ValProAlaLysProGlySerGlyAlaProAlaArgAlaSerCysProG 524
753  GCTACACCCAAATATCAATCCCGCTCTTCAGCCAAATCTTCGCTTC 802
524  yLysIleSerValProValArgAsnLeuHisGlnSerGlyPheSerL 541
803  TAAACACGAGCCCGCCACCATCAGAAACCTCCACCACTTCACCTTCA 852
541  euSerGlySerGlnAlaAspHisIleAlaArgArgThrAlaGlnLys 557
853  TATCAGTCCCGCAGATAGATGACAAATCCAGCGCCACSGG.CACGGC 901
558  IleMetAlaProPro 562
902  ATCTGGGCGCCCGC 916

```

seq_name: gb_est29:AL533854

seq_documentation_block:

LOCUS AL533854 778 bp mRNA EST 13-FEB-2001
 DEFINITION AL533854 LTI_FL013_FBnrl Homo sapiens cDNA clone CS0DF003YA05 5
 prime, mRNA sequence.
 ACCESSION AL533854
 VERSION AL533854.1 GI:12797347
 KEYWORDS EST
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 778)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..778
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF003YA05"
 /clone_lib="LTI_FL013_FBnrl"
 /dev_stage="pooled tissue from post conception fetuses (20
 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by Life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 183 a 232 c 209 g 153 t 1 others
 ORIGIN

alignment_scores:

Quality: 1068.00 Length: 258

Ratio: 4.395 Gaps: 0

Percent Similarity: 94.186 Percent Identity: 77.132

alignment_block:

US-09-367-496-8 x AL533854

Align seg 1/1 to: AL533854 from: 1 to: 778

234 GluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLe 250
 2 GAAGCGGTGAATCGTGCATCACCATGCCAACACAGACCACTGCCGCT 51

250 utyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 52 GTATATCACCAAGGTGATGAGCAAAARCTCTGTGAGGTGATCGCCAGG 101

267 lalysArgGlyValValPheGlyGluProIleThrAlaSerLeu 283
 102 CACGGAAGAGGGAACCTGTGTGTATGCGGAGCCCATCACTGCCGCTTC 151

284 GlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 152 GGAACGAGCGGCTCCCATCTACTGACAGCAAGACTGGGCAAGGCTCTGC 201

300 aPheValThrSerProValAsnProAspProThrThrAlaAspHisL 317
 |||||

www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Insert Length: 711 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 505
POLYA-NO.

POLYA=NO.

FEATURES

```

    .603
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="1048856"
    /clone_lib="Soares_testi
    /sex="male"
    /sex="male"

```

(lab_host="DH10B",
 notc=vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site-1: Not I; Site-2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 Inc., and primed with a Not I - oligo(dT) primer [5',
 GTGTACCAATCTGAAGTGGAGCGCGCCATTTTTTTTTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7T73 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo.
 204 c 200 σ 139 t

BASE COUNT	120 a	204 c	200 g	139 t
ORIGIN				

alignment_scores:		
Quality:	1046.00	Length: 222
Ratio:	4.911	Gaps: 3
Percent similarity:	95.946	Percent Identity: 93.694

alignment_block: y AA778518/rev

Align seg 1/1 to reverse of: AA778518 from: 1 to: 663

seq 1/1

6	AspAsn	pheAlaLeu	lleProGluGlyThr	AsnGlyIleGluGlu	ArgMe	362
13	GACAACTTCGGCGCGAT	CCCCCAGGGCACCAACGC	CAATTCAGGAGCGCAT	614		
62	tSerMet	ValTrpGlu	TyCysValAla	SerGlyLysMet	AspGlu	379
113	GTCCGATGCTCGGAGAAAT	GTGTGGCTCTCGGGAAGAT	GACGAGAATG	564		
79	luPheVal	AlaValThr	SerThrAsnAla	AlaLysIlePhe	AsnPhe	395
63	AGTCTGCTCGCGGTGAC	CAGTACAAATGCTGCCAAAAT	CTTCAATTTTAT	514		
96	ProArgLys	GlyArgVal	AlaValGlySer	AspAlaLeu	ValIle	412
113	CCAAAGAGCGCGAGTGC	GTGTGGGTCTGACCGCTCA	CGCTGTGCATAT	464		
112	rPasnPro	LysAlaThrLysIle	IleSerAlaLysThr	HisAsnLeu	Asn	428
63	GGRACCCCAAGGCCCA	CAAGATCATCTCTGCCAAGAC	CCCAATCTGAAC	414		
29	ValGlu	TyrAsnIlePhe	GluGlyValGluCys	ArgGlyAlaPro	AlaVal	445
1113	GTGGAGTACAACTTC	TGGAGGAGTGGAGTCC	CGGGAGCGCTGCCGT	364		
445	lValIle	SerGlnGlyArg	ValAlaLeuGlu	AspGlyLysMet	PheVal	461
363	GGCTAATGTC	CAGGGCCGAGTGC	GTGGAGCACGGGAAGAT	GTGTGTGTC	314	
462	ThrPro	GlyAlaGlyArg	pheValProArgLysThr	PhePro	AspPheVal	478
313	ACCCCGGGGGTGGCCG	CTTGTCCTCGGAACAAT	TCCCGGACATTCGT	264		
478	lTyrLys	ArgIleLysAla	ArgAsnArgLeu	AlaGluIleHis	GlyVal	494

202 CTTTGTCACTCCCGACCCCTTGAGCCCTGTATCCACCACTCCAGACTTC 251

317 euThrCysLeuLeuSerGlyAspLeuGlnValThrGlySerAlaHis 333

252 TCAACTCCTTGGCTCTCTGTGGAGACCTCCAGGTCAAGGGCAGTGCCTAT 301

334 CysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLe 350

302 TGCACGTTTAACACTGCCAGGAAGGCTGTAGGAAGGACACTTCACCCT 351

350 utlPleProGluGlyThrAsnGlylleGluGluArgMetSerMetValTrpG 367

352 GATTCGGAGGGCCACCAATGGCACCTGAGAGCGGATGTCCGTCATCTGGG 401

367 lulYscYcsValAlaSerGlyLysMetAspGluAsnGlnupPheValAlaVal 383

402 ACAAGCTGTGGTCACTGGGAAGATGGATGAGAACCACTTTGTGGCTGTG 451

384 ThrSerThrAsnAlaAlaLysllePheAsnPheTyrProArgLysGlyAr 400

452 ACCAGCACCAATCAGACCAAGTCTTCAACCTTTTACCCCGGNAAGCCG 501

400 gValAlaValGlySerAspAlaAspLeuVallleTrpAsnProLysAlaTr 417

502 CATTCGTGGGATCCGATGCCACCTGGTCACTGGGACCCCGCAGACGG 551

417 hrLyslleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIle 433

552 TTAAAGACCATCTCTGCCAAGACACAAACAGCTCTCTCGAGTACAACATC 601

434 PheGluGlyValGluCysArgGlyAlaProAlaVallleSerGlnG 450

602 TTTCAAGGCATGGAGTGC CGCGGCTCCCACTGGTGGTCAATCAGCCAGG 651

450 yargValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyA 467

652 GAAGATTTCTCTGGAGGAGCGCACCCCTGCATGTCAACCA. AGCTCTGGAC 700

467 rgpPheValProArgLysThrPheProAspPheValTyrLysArgIleLys 483

701 GCTACATTTCCCGGAAGCCCTTCCCTGTATTTGTTTACACGGTATCAAG 750

484 AlaArgAsnArgLeuAlaGluIle 491

751 CTACAGCAGCAGCTGGCTGAGCTG 774

seq name: qb est11:AA778518

seq_documentation_block: 24-OCT-2000
LOCUS AA778518 663 bp mRNA EST
af59505.1 Soares_testis_NHT Homo sapiens cDNA clone 1048856 3'
DEFINITION similar to TR:000240 000240 CYTOSOLIC PHOSPHOPROTEIN ; mRNA
sequence.

ACCESSION AA778518
VERSION AA778518.1 GI:2837849
KEYWORDS EST.
SOURCE human.

[illegible]

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

JOURNAL
TITLE
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
 Tumor Gene Index
 Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.

CONTACT: Robert Strausberg, Ph.D.
Email: cgaphs@email.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

263 CTCAAGAGGATCAAAAGCTCGCAACAGGCTGCGGAGAGCTCCACGTCGT 214
495 ProArgGlyLeuTyrAspGlyProValHisGluValMetValProAlaLys 511
213 CGCGGTGGGCTGTATGACGGGCGGCTCGACAGAGGTGATGCTGCCTGCCAA 164
511 sProGlySerGlyAlaProAlaAraGlySerCysProGlyLysIleSerV 528
163 GCCAGGAGCTGCGCTCGCGCGCGCGCTGCTGCCAGGCAAGATCTCCG 114
528 alProProValArgAsnLeuHisGlnSerGlyPheSerLeuSerGlySer 544
113 TCCCTCTGTCGGCAACCTACATCAGTCGGGGTTCAGCTATCTGGTCT 64
545 GlnAlaAspAspHisIleAlaArgArgThrAlaGlnLysIleMetAlaPr 561
63 CAGGCTGATGACCATCGCCCGAGGACAGCAGCAGCATCATGCGACC 14
561 oProGlyGly 564
13 CACTGGCGGC 4

seq_name: gb_est22:AI569716

seq_documentation_block:
LOCUS AI569716 612 bp mRNA EST 29-MAR-1999
DEFINITION tn05a06.x1 NCI-CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2166706 3'
similar to TR:014531 O14531 DIHYDROPYRIMIDINASE RELATED PROTEIN 4.
; mRNA sequence.
ACCESSION AI569716
VERSION AI569716.1 GI:4533090
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 612)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 396.
Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2166706"
/tissue_type="NCI-CGAP_Brn25"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT73D-pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCATCTCAAGTGGAGCGCGCATAGGTTTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

```

BASE COUNT 115 a 194 c 172 g 130 t 1 others
ORIGIN

alignment_scores:
Quality: 1029.00 Length: 204
Ratio: 5.145 Gaps: 0
Percent Similarity: 98.039 Percent Identity: 97.549

alignment_block:
US-09-367-496-8 x AI569716/rev
Align seg 1/1 to reverse of: AI569716 from: 1 to: 612

352 ProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGluLys 368
612 CCCGAGTGCACCATCGGCATTGAGGAGCGCATGTCGATGGTCTGGGAGAA 563
368 sCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValThrS 385
562 ATGTGTGGCATCTGGGAAGATGGAGAGATGAGTTCGTCGCGGTGACCA 513
385 erThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArgVal 401
512 GTACAAATGCTGCCAAAATCTTCAATTTTACNCCAAGGAGGCGGAGTG 463
402 AlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThrLy 418
462 GCTGTGGCTCTGACGCTGACCTGTCATATGGAACCCCAAGGCCACAA 413
418 sIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePheG 435
412 GATCATCTGCCAAGACCCACAACTCTGAACGTGGAGTACAAATCTTCG 363
435 luGlyValGluCysArgGlyAlaProAlaValIleSerGlnGlyArg 451
362 AGGAGTGGAGTCCCGGGAGCCCTGCGGTGTCATAAGTCAGGCGCGA 313
452 ValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyArgph 468
312 GTGGCGCTGGAGAACCGGAAGATGTTGTACCCCGGGGCGGCGCTT 263
468 eValProArgLysThrPheProAspPheValTyrLysArgIleLysAlaA 485
262 CGTCCCTCGGAAACATCTCCCGACTTCTGTACAGAGGATCAAGGCTC 213
485 rgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGly 501
212 GCAACAGGCTGGCGAGATCCACGGTGTGCCCGTGGACTGTATGACGG 163
502 ProValHisGluValMetValProAlaLysProGlySerGlyAlaProAl 518
162 CCGCTCCACGAGGTGATGTCCTGCCAAGCCAGGAGTGGCGCTCGCGC 113
518 aArgAlaSerCysProGlyLysIleSerValProValArgAsnLeuH 535
112 CCGCGCTGCTGCCAGGCAAGATCTCGTCCCTCTCTGTCGCGCAACCTAC 63
535 isGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIleAla 551
62 ATCAGTCGGGTTCAGCTATCTGGGTCTCAGGCTGATGACCATGCGC 13
552 ArgArgThrAla 555
12 CGAGCAGCAGCA 1

seq_name: gb_est51:AW71374
seq_documentation_block:
LOCUS AW71374 592 bp mRNA EST 04-MAY-2000
DEFINITION hn55d08.x1 NCI-CGAP_Co17 Homo sapiens cDNA clone IMAGE:3027567 3'
similar to SW:DPY4_HUMAN O14531 DIHYDROPYRIMIDINASE RELATED
PROTEIN-4 ; mRNA sequence.

```

FEATURES
source

us-09-367-496-8.std.ist

Tue Jul 31 13:08:21 2001

469 ValProArgLysThrPheProAspPheValTyrLysArgIleIleysalaar 485
 292 GTCCTCGGAAACATTCCCGGACCTTTGTCTACAGAGGATCAAGCTCG 243
 485 gsnArGLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGlyP 502
 242 CAACAGGCTGGCGGAGATCCACGGGTGCCCCGCTGATGACGGGC 193
 502 roValHisGluValMetValProAlaLysProGlySerGlyAlaProAla 518
 192 CCGTCCACGAGGTATGCTGCTGCCAAGCCAGGAGTGGCGTCCGGCC 143
 519 ArgAlaSerCysProGlyLysIleSerValProValArgAsnLeuH1 535
 142 CCGCGCTGCTGCCAGGCAAGATCTCGTCCCTCTGTGGCAACCTACA 93
 535 sGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIleAlaA 552
 92 TCAGTCGGGGTTCAGCCTATCTGGTCTCAGGCTGATGACCATCGGCC 43
 552 rgArgThrAlaGlnLysIleMetAlaProProGlyGlyArg 565
 42 GACGCACACACAGAGATCATGGCCACCATCTGGCGGCCGC 2

seq_name: gb_est29:AL535160

seq_documentation_block: 1049 bp mRNA EST 13-FEB-2001
 LOCUS AL535160 LTI_FL013.FBrnl Homo sapiens cDNA clone CS0DF008YL11 5
 DEFINITION prime, mRNA sequence.

ACCESSION AL535160
 VERSION AL535160.1 GI:12798653
 KEYWORDS EST
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1049)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source

1..1049
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="LTI_FL013.FBrnl"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 231 a 291 c 328 g 197 t 2 others
 ORIGIN

alignment_scores:
 Quality: 1025.00 Length: 279
 Ratio: 4.116 Gaps: 0
 Percent Similarity: 89.247 Percent Identity: 68.817

AW771374
 AW771374.1 GI:7703429
 EST.
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 592)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps@remail.nih.gov
 Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies
 , Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E.
 Consortium DNA Sequencing by: Washington University Genome
 Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 434.

Location/Qualifiers

1..592
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3027567"
 /clone_lib="NCI-CGAP Col7"
 /tissue_type="juvenile granulosa tumor"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies."
 110 a 187 c 172 g 122 t 1 others

BASE COUNT

alignment_scores:
 Quality: 1025.00 Length: 197
 Ratio: 5.230 Gaps: 0
 Percent Similarity: 99.492 Percent Identity: 99.492

alignment_block:
 US-09-367-496-8 x AW771374/rev ..

Align seg 1/1 to reverse of: AW771374 from: 1 to: 592

369 CysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValThrSe 385
 592 TGTGTGGCTCTGGGAAGATGGAGCAATGAGTTCGCGGGTGACCG 543
 385 rThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArgValA 402
 542 TACAAATGCTGCCAANAATCTTCAATTTTACCAAGGAGGGGAGTGG 493
 402 laValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThrLys 418
 492 CTGTGGGCTCTGACCTGACCTGGTGGTCATATGGAACCCCAAGGCCAAC 443
 419 IleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePheG1 435
 442 ATCATCTCTGCCAAGCAACCAATCTGAACGGTGGAGTACAACATCTTGA 393
 435 uGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGlyArgV 452
 392 GGGAGTGGATGCCGGGGAGCGCTCCCGTGTCTAAGTCAGGGCCGAG 343
 452 alaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyArgPhe 468
 342 TGGCGCTGGAGGACGGGAAGATGTTTGTACCCCCGGGGCGGCCGCTTC 293

alignment_block:

US-09-367-496-8 x AL535160

Align seg 1/1 to: AL535160 from: 1 to: 1049

```

1 MetSerPheGlnGlyLysSerLeuProArgLeuThrSerAspArgLe 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 ATGTCGTACCAAGGCAAGAGAGATCCGCACATCAGGAGTACCGACT 241
17 uLeuLeuArgGlyGlyArgGlyLeuValAsnAspGlnSerPheThrAla 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 CCTCATCAAGGTGGAGGATCATCAAGATGACCAATCCCTTATGCTG 291
34 spValHisValGluAspGlyLeuLeuLeuGlyGlnLeuGluAsnLeuIle 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 ACGTCTACCTGGAGGATGGAGCTTATCAACAAATAGGAGAACTTAATC 341
51 ValProGlyGlyLeuHisThrLeuAspAlaHisCysLeuMetValLeuPr 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 GTTCTCTGGTGGAGTGAAGACCATTTGAAGCCACGGCGGATGTTATCC 391
67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetT 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
392 CGAGGATTTGATGTCAACACCTACCTGCAGAGCCCTCCAGGGGATGA 441
84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGly 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
442 CTGCGCTGATGACTTCTTCAAGGAGGACCGGCGGCTGTTGGCGGG 491
101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
492 ACCAGATGATCATTTGACCATGTTCTTCTGAACCTGGTCCAGCTACT 541
117 uAlaAlaTyrGluGlnTyrArgGluArgAlaAspSerAlaAlaCysCys 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 GACCTCTTTCGAGAGTGGCAGAGGACGACGACCAAAATCTCGTGTG 591
134 spTyrSerLeuHisValAspIleThrArgThrPheHisGluSerIleGlu 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 ATTACTCCTCCAGTGACATGACATCAAGCTGTACGATGGGTTCCGGAG 641
151 GluLeuGluAlaLeuValLysGlyValAsnSerPheLeuValPh 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 GAGCTGGAGTGGTGGTGCAGGCAAGGCGTCAATCTCTTCAAGTCTA 691
167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
692 CATGCCCTATAAGGATGCTACCAAAATGTCCGACAGCCAGCTCTATGAG 741
184 lepHeSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 CCTTTACCTTCTTAAGGGCTGGAGCTGTGATCTGTGCTCCATGCAGAA 791
201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyTl 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 AATGGAGATTGATAGCTCAGGAACAAAGCGGATCCTCGAGATGGCAT 841
217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
842 CACGGTCCCGAGGGCCATGCCTGAGCAGACTGAAGAGCTGGAGGGCG 890
234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
891 AGCGGTGTTCGGGGCCATCACCATTGGGGCGGATCACTCCCTCTGTG 940
251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 TACATCACCAGGTATGAGCAGAGAGTGCAGCCGACATCATCGTCTGCG 990
267 aLysArgArgGlyValValPheGlyGluProfile 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
991 CAGGAAGAAAGGGCCCTAGTTTGTGGAGGCCATTTG 1027

```

seq_name: gb_est29:AL543657

seq_documentation_block:

LOCUS AL543657 1117 bp mRNA EST 16-FEB-2001
 DEFINITION AL543657 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI006YB04 5
 prime, mRNA sequence.
 ACCESSION AL543657
 VERSION AL543657.1 GI:12876136
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1117)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1..1117
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DI006YB04"
 /clone_lib="LTI_NFL006_PL2"
 /tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Teng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : filiang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 275 a 304 c 288 g 246 t 4 others
 ORIGIN

alignment_scores:

Quality: 1015.00 Length: 284
 Ratio: 4.028 Caps: 2
 Percent Similarity: 88.732 Percent Identity: 67.254

alignment_block:

US-09-367-496-8 x AL543657

Align seg 1/1 to: AL543657 from: 1 to: 1117

```

289 HistTyrTrpSerLysAsnTrpAlaLysAlaAlaAlaPheValThrSerPr 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 CATATTGGACAGCAAGACTGGGCAAGCGCTGCTATTTGTGACATCCCC 50
305 oProValAsnProAspProThrThrAlaAspHisLeuThrCysLeuLeuS 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ACCCTGAGCGCTGACCAACTACTCCGGACTACATCACTCTTGTCTGG 100
322 erSerGlyAspLeuGluValThrGlySerAlaHisCysThrPheThrThr 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 CCAGCGGGGATCTGCAGCTATCTGGGAGTCCCACTGCACCTTCAGCACT 150
339 AlaGlnLysAlaValGlyLysAspAsnPheAlaLeuIleProGluGlyTh 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 GCCCAGAAAGCAATTTGGGAAGGACAACTTCACAGCCATTCCTGAGGCGAC 200
355 rAsnGlyIleGluGluArgMetSerMetValTrpGluLysCysValAlas 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 CAATGTTGGAGGAGCGGATGCTGTCTATCTGCGGACAAAGCGTGTGGCCA 250
372 erGlyLysMetAspGluAsnGluPheValAlaValThrSerThrAsnAla 388

```

Tue Jul 31 13:08:21 2001

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10371 row: 1 column: 11
 High quality sequence stop: 690.
 Location/Qualifiers
 1. .697
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4502554"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 169 a 196 c 205 g 127 t
 ORIGIN

FEATURES
source

```

251 CAGGAAATGGAGAAACAGGTCGTGGCTGTGACAAAGCAGCAACAGCT 300
389 AlalysilePheAsnPhetPProAtGLysGLyArgValAlaValGlySe 405
301 GCCAAGATCTTCAACCTGTATCCCGAA. GGAAGAATATCTGTGGTTC 349
405 rAspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleSerA 422
350 TCACAGGCACCTCGTCATCTGGATCCAGATGTGTGAAGATCGTCFCTG 399
422 laLysThrHisAsnLeuAsnValGluTrpAsnIlePheGluGlyValGlu 438
400 CCAAGAACCCAGCTCTGGCAGCAGTACACATCTTTGAAGGATGGAG 449
439 CysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuG1 455
450 CTGCGCGGCTCTCTGTTGTCATCTGCCAGGCGAAGATCATGCTGA 499
455 uAspGlyLysMetPheValThrProGlyAlaGlyArgPheValProArgL 472
500 AGATGGCAACCTGCACGTGACCCAGGGGCTGGCCCTTCATACCTGCA 549
472 yThrPheProAspPheValTyrLysArgIleLysAlaArgAsnArgLeu 488
550 GCCGTTCTCCGACTATGCTACAGGCGCATTAAGCAGCGGAGGATG 599
489 AlaGluIleHisGlyValProArgGlyLeuTyrAspGlyProValHisG1 505
600 GCACACCTGCATGCCCTCCAGGGGCGATGACGATGGGCGCTGTGTGCA 649
505 uValMetValProAlaLysProGlySerGlyAlaProAlaArgAlaSerC 522
650 CTTGACCAACCCCAAGGTGGACCCCGCAGGCTCTGTCTGGGGCT 699
522 ysProGlyLysIleSerValProValArgAsnLeuHisGlnSerGly 538
700 CTCCTACTCGGCGGAAAC...CCACCTGTAAATAATCTTCATCAGTCGGA 746
539 PheSerLeuSerGlySerGlnAlaAspAspHisIleAlaArgArgThrAl 555
747 TTTAGCTGTCTCAGGACTCAAGTGGATGAGGGGTT...CGCTCAGCCAG 793
555 aGlnLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLys 572
794 CAAGCGCATCTGGCGCCCGCCAGCGCGCTTCTTAATATCATCATCTGA 843
572 er 572
844 GT 845

```

seq_name: gb_est98:BG293352

```

seq_documentation_block: 697 bp mRNA EST 21-FEB-2001
LOCUS BG293352 697 bp mRNA EST 21-FEB-2001
DEFINITION 602390746F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502554 5',
mRNA sequence.
ACCESSION BG293352
VERSION BG293352.1 GI:13052953
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

alignment_scores:
  Quality: 995.00 Length: 235
  Ratio: 4.482 Gaps: 3
Percent Similarity: 94.468 Percent Identity: 88.511

alignment_block:
US-09-367-496-8 x BG293352 ..
Align seg 1/1 to: BG293352 from: 1 to: 697

119 AlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAspTy 135
:::|||||
2 AGCTATGAGCAGTGGCGGAGCGAGCAGACAGCGCGC.TGCTGTGACTA 50

135 rSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGluGluL 152
|||:::
51 CTCCTTACATGTGACATCTCTCGCTGGCAGCAGAGCACCACCAAGAGAGC 100

152 euGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPheMet 168
|||||
101 TGGAGGCCCTTACTCAGGACAAAGGTGTGAACCTCTTCTGGTCTTCATG 150

169 AlaTyrLysAspArgCysGlnCysSerAspSerGlnMetIleGluIlePh 185
|||||
151 GCATCAAGGACAGGTGCCAGTGTACTGACGCTCAGATATATGAATCTT 200

185 eSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnG 202
|||||
201 CA3CCTCATCCCGGACCTGGGAGCTGTGGCCAGGTGCAGCGCAGAAAATG 250

202 LyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIleThr 218
|||||
251 GGSACATCTGTGGAGGAGGAACAAGCGCCTGCTGGAGCAAGGCATCACT 300

219 GlyProGluGlyHisValLeuSerHisProGluGluValGluAlaGluAl 235
|||||
301 GGTCTTGAGGGCCATGTGCTCAGCCACCCAGAGAGGTAGAGGCCGAGGC 350

235 aValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyrV 252
|||||
351 TGTGTACAGACGACGTCCACATTTCCCAAGCAGGCGCAACTGCCACCTATAG 400

252 alThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLys 268
|||||
401 TCACCAAGGTGTAGCAGGAGGTGTGAGCTGACATGTTGCTCCCAAGCAAG 450

269 ArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuGlyTh 285
|||||
451 GCGAGGGGGTGGTGGTCTTTTGGGGAACCTATCATCTGCCAGCCTGGGCAC 500

```


285 rAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaLaphev 302
 501 TGATGGCTCACACTACTGAGCAAGAACTGGCCAAAGGCTGACCCCTTG 550
 302 alThrSerProValAsnProAspProThrAla.AspHisLeuTh 318
 551 TCACCTTACCCCTTATCAACCCGGAC...CTACTACTGAGACCACTTAC 597
 318 rCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCyst 335
 598 CTCCTGCTGCTCCAGTGGGACCTCCAGGTCACAGGCACTGCC.CACTGCA 646
 335 hrPheThrAlaGlnLysAlaValGlyLysAspAsnPhaLeuLeu 351
 647 CCTTCAT.ACTGCCCAAGAGCTGTGGCAAAG...ACAATTCACATGATT 692
 352 Pro 352
 693 CCC 595

seq_name: gb_est30:AUI18714

seq_documentation_block:
 LOCUS AUI18714 773 bp mRNA EST 19-OCT-2000
 DEFINITION AUI18714 HEMBA1 Homo sapiens cDNA clone HEMBA1004213 5', mRNA
 sequence.
 ACCESSION AUI18714
 VERSION AUI18714.1 GI:10933849
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawal,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5' & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute
 Location/Qualifiers
 1..773
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1004213"
 /clone_lib="HEMBA1"
 /tissue_type="whole embryo, mainly head"
 /dev_stage="embryo, 10 weeks"
 /note="vector: pME18SF13"

FEATURES

source
 183 a 213 c 219 g 153 t 5 others
 BASE COUNT
 ORIGIN

alignment_scores:

Quality: 992.50
 Ratio: 4.451
 Percent Similarity: 95.708
 Length: 233
 Gaps: 1
 Percent Identity: 78.112

alignment_block:

US-09-367-496-8 x AUI18714 ..
 Align seg 1/1 to: AUI18714 from: 1 to: 773

181 MetTyrGluIlePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnVa 197
 2 ATTTATGAGTACTGAGTGTGATCCGGGATATTGGCGCATAGCCCAAGT 51
 197 lHisAlaGluAsnGlyAspIleValGluGlnLysArgLeuLeuG 214
 52 CCACGAGAAATGGGACATCATTCACAGAGGACGACAGAGGATCCTGG 101
 214 lueuGlyIleThrGlyProGluGlyHisValIleuSerHisProGluGlu 230
 102 ATCTGGGCATCACAGGCCCGGAGGACATGTGCTGAGCGACCTGAGGAG 151
 231 ValGluAlaGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAs 247
 152 GTTCGAGGCGGAAGCGTGAATCGTCCCATCACCATCGCCAAACCCAGAC 201
 247 nCysProLeuTyrValThrLysValMetSerLysGlyAlaAlaAspAla 264
 202 CTGCCCCGTGTATATCACCAAGGTGATGAGCAAAAGCTCTGCTGAGGTCA 251
 264 leAlaGlnAlaLysArgGlyValValPheGlyGluProIleThr 280
 252 TCGCCCGAGCGGAGGAAAGGAACTGTGGTGTATGGCGAGCCCATCACT 301
 281 AlaSerLeuGlyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLy 297
 302 GCCAGCTTGGGAGCGGCGCTCCCATTTACTGGAGCAAGAACTGGGCCAA 351
 297 sAlaAlaAlaPheValThrSerProValAsnProAspProThrThrA 314
 352 GGCTGCTGCTTGTCTACCTCCCGACCTTGGAGCCCTGATCCAAACCACTC 401
 314 laAspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnValThrGly 330
 402 CACACTTTCACACTCCCTGCTGCTGCTGGAGACCTNCAGGTCACGGGC 451
 331 SerAlaHisCysThrPheThrThrAlaGlnLysAlaValGlyLysAspAs 347
 452 AGTGCCCATTTGACGTTTACACTGCCAGAGGCTGTAGGAAAGGACAA 501
 347 nPheAlaLeuIleProGluGlyThrAsnGlyIleGluGluArgMetSerM 364
 502 CTTACCCCTGATTCGGAGGGGACCAATGGCACTGAGGAGCGGATGTCCG 551
 364 etValTrpGluLysCysValAlaSerGlyLysMetAspGluAsnGluPhe 380
 552 TCATCTGGGACAGGCTGTGGTCACTGGGAGATGGATGAGAACAGTTT 601
 381 ValAlaValThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProAr 397
 602 GTGGCTGTGACCAAGCAATGCAGNCAAGTCTTCAACCTTTACCCCG 651
 397 glysGlyArgValAlaValGlySerAspAlaAspLeuValIleTrpAsn 413
 652 GAAAGCCCGCATGTCTGGGATCGATGCCGAC.....TGGCAT 691

seq_name: gb_est29:AL537474

seq_documentation_block:
 LOCUS AL537474 784 bp mRNA EST 13-FEB-2001
 DEFINITION AL537474 LTI_FLO13_Fbrnl Homo sapiens cDNA clone CS0DF025YN09 5
 prime, mRNA sequence.

ACCESSION AL537474
 VERSION AL537474.1 GI:12800967
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)

us-09-367-496-8.std.rst

'Tue' Jul 31 13:08:21 2001

COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. 784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LRI_F1013_FBrnl"
/dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and Eco RV sites of the cDNA and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 188 a 236 c 217 g 143 t

ORIGIN

alignment_scores:
Quality: 985.00 Length: 261
Ratio: 4.246 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 68.582

alignment_block:

US-09-367-496-8 x AL537474 ..

Align seg 1/1 to: AL537474 from: 1 to: 784

```

277 GluProTleThraSerLeuGlyThrAspGlySerHisTyrTrpSerly 293
|||||
2 GAGCCATTGCCCGCAGCGCTGGGACCGATGGCACCATTACTGAGCAA 51
|||||
293 sAsnTtpAlaLysAlaAlaPheValThrSerProProValAsnProA 310
|||||
52 GAACCTGGGCCAAGGCTGCGCGCTGCTGACCTCCCTCCCTCCCTCC 101
|||||
310 spProThrAlaAspHisLeuThrCysLeuLeuSerSerGlyAspLeu 326
|||||
102 ACCCTACACCGCCGCGCTGCTGACCTCCCTCCCTCCCTCCCTCC 151
|||||
327 GlnValThrGlySerAlaHisCysThrPheThrThrAlaGlnLysAla 343
|||||
152 CAGGTCACAGGCGCGCCACTGTCTTACAGCAGCTGCCCGGAGCGGT 201
|||||
343 lGlyLysAspAsnPhelAlaLeuileProGluGlyThrAsnGlyileG 360
|||||
202 GGGCAAGGAGCAACTTTACCTGATCCCGAGGCTGTCACGGGATAGG 251
|||||
360 luArgMetSerMetValTrpGluLysCysValAlaSerGlyLysMetAs 376
|||||
252 ACCGGATGCGCTCTGCGGACAGCGGTGGCTACTGGCAAAATGGAT 301
|||||
377 GluAsnGluPheValAlaValThrSerThrAsnAlaLysilePheAs 393
|||||
302 GAGAACCATGTTTGTGCTGCTACAGCAGCAATGCGAGCAAGATCTTAA 351
|||||
393 nPheTyrProArgLysGlyArgValAlaValGlySerAspAlaaspLeu 410
|||||
352 CCGTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 401
|||||
410 alletRpsnProLysAlaThrLysileileSerAlalysThrHisAsn 426
|||||
402 TCATCTGGGACCCCGACAGATTGAAGACCATTAACAGCCAAAGTCA 451

```

```

427 LeuAsnValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaPr 443
|||||
452 TCGCGCGTGGAGTACAAACATCTTCGAGGGTATGGATGCCAGCGCTCCC 501
|||||
443 oAlaValValIleSerGluGlyArgValAlaLeuGluAspGlyLysMetP 460
|||||
502 ACTACTGCTGTCATCAGCCAGGCGCAAGATCGCTTTGAAGACGGAACATCA 551
|||||
460 heValThrProGlyAlaGlyArgPheValProArgLysThrPheProAsp 476
|||||
552 ACGTCAACAGGCGATGGCGCGCTTCATTCGCGGAAGCGTTCCCGGAG 601
|||||
477 PheValTyrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisG 493
|||||
602 CACCTGTACAGCGCGCTCAAAATCAGGAATAGGTTTGGATTGCAAGG 651
|||||
493 yValProArgGlyLeuTyrAspGlyProValHisGluValMetValProA 510
|||||
652 GGTTCAGGCGCGATGATGACGGTCTCTGTAGAGGTACCGAGCTACAC 701
|||||
510 laLysProGlySerGlyAlaProAlaArgAlaSerCysProGlyLysile 526
|||||
702 CCAATATGCAACTCCCGCTCTTCAGCCAAATCTTCGCTTCTTAAACAC 751
|||||
527 SerValProProValArgAsnLeuHisGlnSer 537
|||||
752 CAGCCCCCACCACCATCAGAAACCTCCACCAAGTCC 784

```


.Tue Jul 31 13:08:21 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:38:11 ; Search time 23.39 Seconds
(without alignments)

3235.506 Million cell updates/sec

Title: US-09-367-496-8

Perfect score: 2985

Sequence: 1 MSFQGGKSIPTSDRLIR.....RTAKIMAPGGRSNTLSLS 572

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL16.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.unclassified.*
- 13: sp.vertebrate.*
- 14: sp.virus.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2517	84.3	574	13 Q9DDZ6	Q9ddz6 gallus gall
2	1511	50.6	519	11 Q9E0F5	Q9eqf5 mus musculus
3	1475.5	49.4	564	11 Q9JHU0	Q9jhu0 rattus norv
4	1475.5	49.4	564	11 Q9E0F6	Q9eqf6 mus musculus
5	1458.5	48.9	564	4 Q9NC4	Q9nc4 homo sapien
6	1453.5	48.7	564	4 Q9NRY9	Q9nry9 homo sapien
7	1450	48.6	563	11 Q9JMG8	Q9jmg8 rattus norv
8	1286.5	43.1	589	5 Q9V3N7	Q9v3n7 drosophila
9	1074.5	36.0	479	2 Q9I676	Q9i676 pseudomonas
10	1043	34.9	531	10 Q9FMP3	Q9fmp3 arabidopsis
11	834.5	28.0	332	5 Q9VND9	Q9vnd9 drosophila
12	692.5	23.2	467	2 Q9E809	Q9e809 streptomyce
13	523.5	17.5	542	3 Q9P903	Q9p903 saccharomyce
14	402.5	13.5	458	2 Q9F465	Q9f465 artibacille
15	317	10.6	438	2 Q9KAH8	Q9kah8 bacillus ha
16	282	9.4	430	2 Q9EV52	Q9ev52 staphylococ
17	259.5	8.7	422	1 Q9HMH9	Q9hnh9 halobacteri
18	240.5	8.1	428	2 Q9KX83	Q9kxr3 streptomyce
19	233.5	7.8	428	2 Q9K9V7	Q9k9v7 bacillus ha

20	231.5	7.8	2198	5	Q18990	O18990 caenorhabdi
21	230.5	7.7	2188	5	Q9VXD5	Q9vxd5 drosophila
22	213	7.1	449	2	Q9PEP0	Q9pep0 xylella fas
23	203	6.8	708	11	Q60504	Q60504 cricetus
24	200	6.7	445	2	Q9HT33	Q9ht33 pseudomonas
25	199	6.7	423	2	Q9F415	Q9f415 pseudomonas
26	199	6.7	499	5	Q9VWM1	Q9vwm1 drosophila
27	192	6.4	474	11	Q35102	Q35102 rattus norv
28	184	6.2	414	1	Q9HIM0	Q9him0 thermoplasma
29	183.5	6.1	409	2	Q50432	Q50432 mycobacteri
30	172	5.8	2275	3	Q93937	Q93937 emericella
31	161	5.4	569	1	Q74030	Q74030 methanobact
32	158.5	5.3	569	1	Q9578	Q9578 methanobact
33	158	5.3	397	2	Q9KQ07	Q9kq07 bacillus ha
34	158	5.3	569	1	Q27599	Q27599 methanobact
35	157.5	5.3	752	5	Q97163	Q97163 drosophila
36	155	5.2	415	2	Q9SLC6	Q9slc6 arthrobacte
37	154	5.2	404	5	Q9NL79	Q9nl79 leishmania
38	153.5	5.1	381	2	Q9KXV7	Q9kxv7 streptomyce
39	153	5.1	431	1	Q9HN51	Q9hn51 halobacteri
40	144.5	4.8	386	2	Q9X247	Q9x247 thermotoga
41	144.5	4.8	402	2	Q9KSO1	Q9ksq1 vibrio chol
42	140.5	4.7	526	1	Q9V2D3	Q9v2d3 pyrococcus
43	136	4.6	569	2	Q9S005	Q9s005 helicobacte
44	134.5	4.5	383	5	Q76138	Q76138 trypanosoma
45	131.5	4.4	445	2	Q9KEV3	Q9kev3 bacillus ha

ALIGNMENTS

RESULT 1
Q9DDZ6 PRELIMINARY: PRT: 574 AA.
AC Q9DDZ6
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE COLLAPSEIN RESPONSE MEDIATOR PROTEIN 3.
GN CRMP3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20545548; PubMed=10956643;
RA Fukuda M., Watakabe I., Yuasa-kawada J., Kawachi H., Kuroiwa A.,
RA Matsuda Y., Noda M.;
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
Response Mediator Protein Family,"
RL J. Biol. Chem. 275:37957-37965(2000).
DR EMBL; AF249294; AAG37997.1; -;
SQ SEQUENCE 574 AA; 62070 MW; 96E17AC0E96CA5F4 CRC64;

Query Match 84.3% Score 2517; DB 13; Length 574;
Best Local Similarity 81.4%; Pred. No. 1e-173;
Matches 467; Conservative 56; Mismatches 49; Indels 2; Gaps 1;

QY	1	MSFQGGKSIPTSDRLIRGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA	60
DB	1	MSNPGRSGPRCTSEQLLKIGKIVNDQSFYADYIYEDGLIKOIGENLAVPSGVRVDA	60
QY	61	HGLMVLPGVDVHTRLQPLVLTGTPADFCGCTKAALAGGTTMILDRVFPDTPGVSLLAA	120
DB	61	YGLVYVPGGIDVHTRLQPLVLTGTPADFCGCTKAALAGGTTMILDRVFPDTPGVSLLAA	120
QY	121	EQWRERADSNACCDYSLHVDITRWHESSKEELEALVKEGVNSFLVFMAYKRCQCSDSQ	180
DB	121	EQMRQRADSNACCDYSLHVDITRWHESSKEELEALVKEGVNSFLVFMAYKRCQCSDSQ	180

QY 181 MYEIPSIIRDLGALAOVHAENGDIVBEEQKRLLELIGITGPEGHVLSHPBEVEAEVYRAV 240
 DB 181 MYEIPSIIRDLGALAOVHAENGDIIBEEQKRLDITGPEGHVLSRPEVEAEVYRAI 240
 QY 241 TIAQANCPPLYVTKVMSKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAA 300
 DB 241 TIAQANCPPLYVTKVMSKGAADVAAQAKRGTVVYGEPTASLGADGSHYWSKNWAKAAA 300
 QY 301 FVTSPPVNDPTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTIEE 360
 DB 301 FVTSPPVNDPTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
 DB 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
 QY 421 SAKTHNLNVEYIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFVYK 480
 DB 421 SAKTHNLNVEYIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMYPAK--PGSGAPARASCPGKISVPPVRLHQSG 538
 DB 481 RIKARNRLAEIHGVRGLYDGPVHEVMYPAK--PGSGAPARASCPGKISVPPVRLHQSG 538
 QY 539 FSLSGQADHDHARTAKOIKINAPPGGRSNTISLS 572
 DB 541 FSLSGQADHDHARTAKOIKINAPPGGRSNTISLS 574

RESULT 2

Q9EQF5 ID Q9EQF5 PRELIMINARY; PRT; 519 AA.
 AC Q9EQF5;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE DIHYDROPRIMIDINASE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER/NIH; TISSUE=WHOLE EMBRYO;
 RX MEDLINE=20545548; PubMed=10956643;
 RA Fukuda M., Watakabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,
 Matsuda Y., Noda M.;
 RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
 Response Mediator Protein Family."; J. Biol. Chem. 275:37957-37963(2000).
 DR EMBL: AF249296; AAG37999.1; -.
 SQ SEQUENCE 519 AA; 56681 MW; F2C53836B2120ABD CRC64;

Query Match 50.6%; Score 1511; DB 11; Length 519;
 Best Local Similarity 58.3%; Pred. No. 4.7e-101;
 Matches 287; Conservative 66; Mismatches 135; Indels 4; Gaps 1;

QY 16 RLIRGRVNDQSFYADVHVEDGLIKOIGENLIVP----GGIHTTDAHGLMVLPGVDV 71
 DB 6 RLIRGRVNDQSFYADVHVEDGLIKOIGENLIVP----GGIHTTDAHGLMVLPGVDV 71
 QY 72 VTRLOMPVLGMPADDCQCGTAAAGTMTLDHVPDPTGVSLAAAYEQRERADSA 131
 DB 66 THTHMQFFMGSQSDVDFYQGTAAAGTMTLDHVPDPTGVSLAAAYEQRERADSA 131
 QY 132 CCDSLVHVDITRWHESTKEELALVKEGVNSFLVFMAYDKRCQSDSQMVEIFSIIRDL 191
 DB 126 CCDSLVHVDITRWHESTKEELALVKEGVNSFLVFMAYDKRCQSDSQMVEIFSIIRDL 191
 QY 192 GALAOVHAENGDIVBEEQKRLLELIGITGPEGHVLSHPBEVEAEVYRAVITAKQANCP 251
 DB 192 GALAOVHAENGDIVBEEQKRLLELIGITGPEGHVLSHPBEVEAEVYRAVITAKQANCP 251

DB 186 GAIAOVHAENGDLIAEGAKKMLALGITGPEGHVLSHPBEVEAEVYRAVITAKQANCP 245
 QY 252 VTKVMSKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAAFTSPVNDP 311
 DB 246 VVHVMSKAAVADARRAGNVYGEPTAAGLGTGQYWESESHAAHHVWGPPLRDP 305
 QY 312 TTADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTIEE 371
 DB 306 LTPGFLMDLLANGDLTTGSDNCTFNTCQKALGKDDFTKIPNGVNGVEDRMSVLEWKG 365
 QY 372 SGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKIIISAKTHNLNVEY 431
 DB 366 SGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKIIISAKTHNLNVEY 431
 QY 432 NIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFVYKRIKARNRLAEI 491
 DB 426 NIFGVEGRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDFVYKRIKARNRLAEI 491
 QY 492 HGVPRLYDGPV 503
 DB 486 VPKRAPHYKGEV 497

RESULT 3

Q9JHU0 ID Q9JHU0 PRELIMINARY; PRT; 564 AA.
 AC Q9JHU0;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE ULIP-LIKE PROTEIN
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Horiuchi M., Betz H.;
 RT "Ulup and dihydropyrimidinase like protein."; Submitted (DSC-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AJ31436; CAB95193.1; -.
 DR InterPro; IPR002195; -.
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61540 MW; E078B9002F54975E CRC64;

Query Match 49.4%; Score 1475.5; DB 11; Length 564;
 Best Local Similarity 50.2%; Pred. No. 2e-98;
 Matches 281; Conservative 95; Mismatches 173; Indels 11; Gaps 3;

QY 14 SDRLLIRGRVNDQSFYADVHVEDGLIKOIGENLIVP---GGIHTTDAHGLMVLPGVDV 73
 DB 7 SVRLIKGGKVVNDCTHEADVYIENGIIQVGRLEMPGGAKVIDATGKLVIPGIDTS 66
 QY 74 TRLOMPVLGMPADDCQCGTAAAGTMTLDHVPDPTGVSLAAAYEQRERADSA 133
 DB 67 THTHMQFFMGSQSDVDFYQGTAAAGTMTLDHVPDPTGVSLAAAYEQRERADSA 133
 QY 134 DYSLVHVDITRWHESTKEELALVKEGVNSFLVFMAYDKRCQSDSQMVEIFSIIRDLGA 193
 DB 127 DYALHVGITWAPFKVKAEMETLVREKGVNSFQNFMTYKDLMLRSELYQVHFACRDFGA 186
 QY 194 LAOVHAENGDIVBEEQKRLLELIGITGPEGHVLSHPBEVEAEVYRAVITAKQANCP 253
 DB 187 IPRVHAENGSELVAEAKAEALDITGITEGIEISHPELEAEATHRVITIANTHCPILV 246
 QY 254 KVMKGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAAFTSPVNDP 313
 DB 247 NVSSISAGDVIAAAMQCKVYLAETNAHAILTGUYTHQDMSHAAAYVTPPLRLDNT 306
 QY 314 ADHLITCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTIEE 373
 DB 307 STYLSLLANDTLNIVASHRPFITTKOKMGKEDFTKIPHGVSQVQDRMSVWGVGG 366

```

Qy 374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKIISAKTHNLNVEYNI 433
Db 367 KMDENRFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTOVQGGDFNL 426
Qy 434 FEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPFVYKRIKARNRLAEIHG 493
Db 427 YENMRCHGVPLVITSRGVYENGVMCAEGTKGFCPLRSFPDIVYKLVQREKTLKVRG 486
Qy 494 VPRGLYDGPVHEVMVPAKPGSGAPARASCPCPKISVPPV-----RNHQSFGSLSGSQAD 547
Db 487 VDRTPYLDGVAIVNPGKEMGTPL-ADTPTR----PVTRHGMGRDLHESFSLSGSQID 541
Qy 548 DHIAARTAAQIMAPPGGRSN 567
Db 542 DHVPRASARILAPPGGRSS 561

RESULT 4
Q9EQF6 PRELIMINARY; PRT; 564 AA.
AC Q9EQF6 01-OCT-2000 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE COLLAPSED RESPONSE MEDIATOR PROTEIN 5.
GN CRMP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/NIB; TISSUE=WHOLE EMBRYO;
RX MEDLINE=20345548; PubMed=10956643;
RA Fukuda M., Watakabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,
RA Matsuda Y., Noda M.;
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
RT Response Mediator Protein Family.";
RL J. Biol. Chem. 275:37957-37965(2000).
DR EMBL: AF249295; AAG37998.1;
SQ SEQUENCE 564 AA; 61516 MW; CA93790FC8F9CD98 CRC64;

Query Match 49.4%; Score 1475.5; DB 11; Length 564;
Best Local Similarity 50.2%; Pred. No. 2e-98;
Matches 281; Conservative 95; Mismatches 173; Indels 11; Gaps 3;

Qy 14 SDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAGLMVLPGGVDVH 73
Db 7 SVRLIKGGKVVNDCTHEADVYIESGIIQVGRGLMIPGAKVIDATGKLVIPGGIDTS 66
Qy 74 TRLQMPVLGMPADDFCOGTKAALAGGTTMILDHVFPDGTGVSLLAAVEQWRERADSACC 133
Db 67 THFHQTFMNAATCVDVDFYHGTGAALVGGTTMILGHVLPDKETSLVAYEKRCALADPKVCC 126
Qy 134 DYSLHVDITRWHEISKEELEALYKEGVNSFLVPMAYKDRCCSDSOMYEIESTIRDLGA 193
Db 127 DYALHVGITWWAPKVAEMETLVREKGVNSFQMFMTYKDLMLRDSLEYQVHFACRDIGA 186
Qy 194 LAQVHAENGDIIVEEQRLLELGTGPEGHVLSHPPEVEAEAVYAVYTIKQANCPLYVT 253
Db 187 IPRVHAENGELVAEGAKALDLGITGPEGIEISHPEELEAEATHRVITIANRTHCPILYV 246
Qy 254 KVMKGAADAIAQAKRRGVVFGEPITASIGTDSHSHVSKNKAFAAFVTSPPVNPDPPT 313
Db 247 NVSSISAGDVIAAKMKGKVVLAETTNHAATLTGLHYHODMSHAAAYVTPPLRLDNT 306
Qy 314 ADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTEERMSVMWKECVASG 373
Db 307 STYLSLLANDTLNIVASDRHPTTKQKAMGKEDFTKPHGVSGVQDRMSVWVGVRGG 366
Qy 374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKIISAKTHNLNVEYNI 433

```

```

Db 367 KMDENRFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTOVQGGDFNL 426
Qy 434 FEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPFVYKRIKARNRLAEIHG 493
Db 427 YENMRCHGVPLVITSRGVYENGVMCAEGTKGFCPLRSFPDIVYKLVQREKTLKVRG 486
Qy 494 VPRGLYDGPVHEVMVPAKPGSGAPARASCPCPKISVPPV-----RNHQSFGSLSGSQAD 547
Db 487 VDRTPYLDGVAIVNPGKEMGTPL-ADTPTR----PVTRHGMGRDLHESFSLSGSQID 541
Qy 548 DHIAARTAAQIMAPPGGRSN 567
Db 542 DHVPRASARILAPPGGRSS 561

RESULT 5
Q9NOCA PRELIMINARY; PRT; 564 AA.
AC Q9NOCA 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RA Horiuchi M., Betz H.;
RT "Human homologue of Ulip6.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ251275; CAB95124.1;
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
KW Hypothetical protein.
SQ SEQUENCE 564 AA; 61394 MW; F370DD35B0B49935 CRC64;

Query Match 48.9%; Score 1458.5; DB 4; Length 564;
Best Local Similarity 49.5%; Pred. No. 3.3e-97;
Matches 277; Conservative 100; Mismatches 172; Indels 11; Gaps 3;

Qy 14 SDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAGLMVLPGGVDVH 73
Db 7 SVRLIKGGKVVNDCTHEADVYIESGIIQVGRGLMIPGAKVIDATGKLVIPGGIDTS 66
Qy 74 TRLQMPVLGMPADDFCOGTKAALAGGTTMILDHVFPDGTGVSLLAAVEQWRERADSACC 133
Db 67 THFHQTFMNAATCVDVDFYHGTGAALVGGTTMILGHVLPDKETSLVAYEKRCALADPKVCC 126
Qy 134 DYSLHVDITRWHEISKEELEALYKEGVNSFLVPMAYKDRCCSDSOMYEIESTIRDLGA 193
Db 127 DYALHVGITWWAPKVAEMETLVREKGVNSFQMFMTYKDLMLRDSLEYQVHFACRDIGA 186
Qy 194 LAQVHAENGDIIVEEQRLLELGTGPEGHVLSHPPEVEAEAVYAVYTIKQANCPLYVT 253
Db 187 IARVHAENGELVAEGAKALDLGITGPEGIEISRPEELEAEATHRVITIANRTHCPILYV 246
Qy 254 KVMKGAADAIAQAKRRGVVFGEPITASIGTDSHSHVSKNKAFAAFVTSPPVNPDPPT 313
Db 247 NVSSISAGDVIAAKMKGKVVLAETTNHAATLTGLHYHODMSHAAAYVTPPLRLDNT 306
Qy 314 ADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGTEERMSVMWKECVASG 373
Db 307 STYLSLLANDTLNIVASDRHPTTKQKAMGKEDFTKPHGVSGVQDRMSVWVGVRGG 366
Qy 374 KMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKIISAKTHNLNVEYNI 433
Db 367 KMDENRFVAVTSSNAKILNLYPRKGRITPGADADVVDVWDPEATKTIISASTOVQGGDFNL 426

```

Qy	434	FEGVCEGCAVAVISQGRVALEDGMEVTPGAGRFVPRKTFPDVYVKRIKARNRLAEIHG	493
Db	427	YENMRCHGVLPTISRGVVYGVVENCAGETGKFCPLRSFPDVTYVKLVQREKTKLVKG	486
Qy	494	VPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPV	547
Db	487	VDRPYLGDAVVVHPGKKEMGTPL-ADTPTP----	541
Qy	548	DHIARRTAQKIMAPPGGRSN	567
Db	542	DHPVKRASARILAPPGGRSS	561

RESULT 6

Q9NRY9	PRELIMINARY;	PRT;	564 AA.
ID	Q9NRY9		
AC	Q9NRY9;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DE	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	COLLAPSN RESPONSE MEDIATOR PROTEIN-5.		
OS	CRMP5.		
GN	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;			
RP	[1]		
RN	SEQUENCE FROM N.A.		
RT	Yu Z., Kiyzer T.J., Griesmann G.E., Lennon V.A.;		
RT	"collapsin response mediator protein-5 (CRMP5), a novel autoantigen		
RT	associated with paraneoplastic neurological disorders,"		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF157634; AAF80348.1;		
DR	InterPro; IPR002195;		
DR	Pfam; PF00744; Dihydroorotase; 1.		
SEQUENCE	564 AA; 61459 MW; 949997A983735340 CRC64;		

[illegible]

QY	494	VPRLGYDGPVHEVMVPAKPGSGAPARASCPGKISVPPV-----RNLHQSGLSLGSGQAD	547
Db	487	VDRTPYLGDVAVVVHPCKKEMGTPL-ADTPTR---PVTRHGGMRDLHSSFSLSGSGQID	541
QY	548	DHIAERTAAOKINAPPGGRSN	567
Db	542	DHVPKRASAKILAPPGGRSS	561
<p>RESULT 7</p> <p>ID Q9JMG8 PRELIMINARY; PRT; 563 AA.</p> <p>AC Q9JMG8;</p> <p>DT Q1-OCT-2000 (TrEMBLrel. 15; Created)</p> <p>DT Q1-OCT-2000 (TrEMBLrel. 15; Last sequence update)</p> <p>DT Q1-NAR-2001 (TrEMBLrel. 16; Last annotation update)</p> <p>DE DIHYDROPYRIMIDINASE-RELATED PROTEIN.</p> <p>OS Rattus norvegicus (Rat).</p> <p>OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;</p> <p>OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;</p> <p>OX NCBI_TaxID=10116;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RA Yanagi S., Inatome R., Yamamura H.;</p> <p>RT "Dihydropyrimidinase-related protein".</p> <p>RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.</p> <p>DR EMBL; AB029432; BAB89475.1; -.</p> <p>DR InterPro; IPR002195; -.</p> <p>DR Pfam; PF00744; Dihydroorotase; 1.</p> <p>SQ SEQUENCE 563 AA; 61395 MW; B36767BF1440004D CRC64;</p>			
<p>Query Match 48.6%; Score 1450; DB 11; Length 563;</p> <p>Best Local Similarity 49.88; Pred. No. 1.4e-96;</p> <p>Matches 279; Conservative 94; Mismatches 175; Indels 12; Gaps</p>			
QY	14	SDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAGLMVLPQGVDPVH	73
Db	7	SVRLIRGGKVNDCTHEADAVIENGIIQVGRLEMPGAKVDTATGKLVPGCIDTS	66
QY	74	TRLQMPVLGHTPADDCCGTPKALAGGTTMLDHPVDPDGVSLLAAYQWRRERADSAACC	133
Db	67	THFQTFNATCVDDFVHGTKAALVGTTMLIGHVLPKETSLAVEYKCRALADPKV-C	125
QY	134	DYSLHVDITRWHSIKEEALVKRGNVSLFVFWAYKDRCCDSQWEIFSIIRDIGA	193
Db	126	DYALHVGITWAPKKAEMETLVREKGVNSFQMFYTKDLYMLRDSLEYQVPHACRDFGA	185
QY	194	LAOVHAENGDIVVEEOKRLLLEGITGPEGHVLSHPEEVAEAVYAVVTIAQANCPLYVT	253
Db	186	IPRVHAENGELVAEGAKALDGLITGPGIEISHPPEELEAEATHRVITIANRTHCPTLV	245
QY	254	KVMSKGAADAIAQAKRRGVVFEPTISLGTGDSHYKSNKWAFAAFVTSPPVPPDPTT	313
Db	246	NVSSISAGDVIAAKMQGVKLVAETTNHAHILTLGLHYHQDWSHAAAYVTPPLRLDNT	305
QY	314	ADHLTCLLSGDLQVTSACHTFTTAQKAVGKONFALIPGCTNGIEBMSVMWEKCVASG	373
Db	306	STYLMSLLANDTLNIVASDRHPTFTKQAMGKEFTKPHGVSGVQDRMSVMWVERGVVG	365
QY	374	KMDNEFVAVTSTNAKIFNFPKRGVAVGSDADLVNPNKATKIISAKTHNLNVEYNI	433
Db	366	KMDNEFVAVTSSNAKIINLYPKRGIIPGADADVVDPEATKTIISASTQVGGDFNL	425
QY	434	FEGVECRGAPAVVISQGRVALDEGKMFVTPGAGRFVPKTFPDPVYKRIKARNRLAEIHG	493
Db	426	YENMRCHGVPLVTISRGVAVVYENGVMCAEGTGNFCPLRSPDVIYKLVQREKTLKVRG	485
QY	494	VPRLGYDGPVHEVMVPAKPGSGAPARASCPCKISVPPV-----RNLHQSGLSLGSGQAD	547
Db	486	VDRTPYLGDVAVVNPGEKEMGTPL-ADTPTR---PVTRHGGMRDLHSSFSLSGSGQID	540

QY	548	DHIARRTAQKIMAPPGRSN	567	DR	EMBL; AF160900; AAD46840.1; -
Db	541	DHVPKRASARILAPPGRSS	560	DR	EMBL; AF077837; AAD52007.1; -
				DR	FlyBase; FBgn0023023; CRMP.
				DR	InterPro; IPR002195; -
				DR	Pfam; PF00744; Dihydroorotase; 1.
				SQ	SEQUENCE: 589 AA; 64430 MW; 959A44B657CF3891 CRC64;
RESULT	8				
Q9V3N7		PRELIMINARY;	PRT; 589 AA.		
ID	Q9V3N7				
AC	Q9V3N7				
DT	01-MAY-2000	(TrEMBLrel. 13, Created)			
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)			
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)			
DE	CRMP PROTEIN				
DE	CRMP OR BCDA:HL02693 OR CG1411.				
GN	Drosophila melanogaster (Fruit fly).				
OS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BERKELEY;				
RA	MEDLINE-20196006; PubMed-10731132;				
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers J., Blazey R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch L., Beasley E.M.,				
RA	Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolshakov S.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dubin K.J., Evangelista C.C., Ferraz C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,				
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,				
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,				
RA	Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,				
RA	Spieler E., Spradling A.C., Turner E., Venter E., Wang A.H., Wang X.,				
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,				
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,				
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,				
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,				
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;				
RT	"the genome sequence of Drosophila melanogaster.";				
RL	Science 287:2185-2195(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,				
RA	Lewis S.E., Suh C., Rubin G.M.;				
RT	"Full Length Drosophila melanogaster cDNA sequence.";				
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Wang L.H., Strittmatter S.M.;				
RT	"Drosophila CRMP is expressed in the developing nervous system.";				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
EMBL	AE003602; AAF52002.1; -				

DR	EMBL; AF160900; AAD46840.1; -
DR	EMBL; AF077837; AAD52007.1; -
DR	FlyBase; FBgn0023023; CRMP.
DR	InterPro; IPR002195; -
DR	Pfam; PF00744; Dihydroorotase; 1.
SQ	SEQUENCE: 589 AA; 64430 MW; 959A44B657CF3891 CRC64;
Query Match	43.1%; Score 1286.5; DB 5; Length 589;
Best Local Similarity	43.9%; Pred. No. 9.7e-85;
Matches	266; Conservative 92; Mismatches 175; Indels 73; Gaps 9;
QY	7 KSP---RITSDBLLRGGRIIVNDOSFYADVHVEDGLIKOIG--ENLIVPGGIHTIDAH 61
Db	9 KKPPIHQSQAQNRYIKNGEIVNHDKSFADVIYIEDGIKIFVGPSSSEITIPGVRITDS 68
QY	62 GLAVLPQGVVHTRLQMPVLGTMTPADDFCOGTAKAALAGGTTMILDHVPDTPGTGSLAAAYE 121
Db	69 GLMIIPGGIDPHTHMQLPFGGAVAVDDFYHGTRKAAVAGGTTMIIDFVLPKHKHESMIEAYD 128
QY	122 QWRERADSAACDYSLVHVDITRWGHESIKKELEALVKEGVNSFLVFMAYKDRQCSDSQM 181
Db	129 KWRWADPKKVCDDYGLHVGITWMSKVSSEIGILCKELGVNSFKTFMAYKGLYQLNDSDL 188
QY	182 YEIFSIIIRDLGALAAQVHAENGDIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRAVT 241
Db	189 LDVFERIHLNGVAMVHAENGDIITAKNTQRLLAEGINGPEGHLSRPEVEAEAVHRAVY 248
QY	242 IAKQANCPLYVTKVMSKGAADAIQAOKRGVVVFGPEPTASLGTD-----GSHWWSKNW 295
Db	249 LAHQMKTPLFVSGLTSSSAELVGRARRSGYCVFGETLASSLGRSMSAVPKGDRIIYA--- 305
QY	296 AKAAAFVTSPPVNDPTADHLTCLLSGDDQVTSQSAHCTTTAQKAVGKDNFALIEPT 355
Db	306 -----ITSPPIRESAETPRQLMKSALVDDQLQGLSDNCTFNKEHKGALGDKFTKIPNGV 359
QY	356 NGIERMSNVWEKCVASGKMDENEFVAVTSTNAKIFNFPKRGVAVGSDADLVINPK 415
Db	360 NGVEDMSLVWEKGVHAGLLDPCRFVAVTSTNAKIFNFPKRGVAVGSDADLVINPN 419
QY	416 AVKIIASAKTNLNVYNIYFEGVECGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTFP 475
Db	420 AVRTISKTHHACDFNIFEGMTVHGVCDFVLVRGICAEGRNVAVAGFGFIFTPVRP 479
QY	476 DVIYKRIKAR-----NRLAEIHGVPRGLYDGPVHVHVMVPAKPGSGAP 517
Db	480 PIVYDILIEKGVQSQPEEQHEKQNGSMARFAEL-----DIQIPVQ-----EP 522
QY	518 ARASCPGKISVPP-----VRLHQSGESLSGSOADDHIAFRTAOKIMAP 561
Db	523 IGAMLAGNLAMPAGSLCSTPSVGRVGRVGRDQLQESSFSI---SEELDRSGVRAKIVKNP 580
QY	562 PGGRSN 567
Db	581 PGKSS 586
RESULT	9
Q91676	
ID	Q91676 PRELIMINARY; PRT; 479 AA.
AC	Q91676;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	DHYDROXYRIMIDINASE.
GN	PA0441.
OS	Pseudomonas aeruginosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC	Pseudomonas.
OX	NCBI_TaxID=287;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-PA01;

[illegible][illegible]

```

RESULT 10
99FMP3
D Q9FMP3 PRELIMINARY; PRT; 531 AA.
O Q9FMP3;
T 01-MAR-2001 (TReMBLrel. 16, Created)
T 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
T 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
E DIHYDROPYRIMIDINASE.
S Arabidopsis thaliana (Mouse-ear cress).
C Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
C Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
C Brassicales; Brassicaceae; Arabidopsids.
X NCBI_TaxID=3702;
X [1]
N SEQUENCE FROM N. A.
P STRAIN=COLUMBIA;
C MEDLINE=98162728; PubMed=9501957;
X Nakamura Y., Sato S., Kaneko T., Kotani H., Miyajima N.,
A Tabata S.:
P "Structural analysis of Arabidopsis thaliana chromosome 5. III.

```

RT Sequence features of the regions of 1,191,918 bp covered by seventeen RT physically assigned pl clones.^a;
 RL DNA Res. 4:401-414(1997).
 DR EMBL: AB007727; BAB10038.1; -.
 SQ SEQUENCE 531 AA: 57991 MW: F6D5F23D73027790 CAC6A.

Query Match	34.9%	Score 1043;	DB 10;	Length 531;
Best Local Similarity	44.4%	Pred. No. 3.3e-67;		
Matches 209; Conservative	84;	Mismatches 166;	Indels 12;	Gaps 5;

[illegible]

RESULT	11
ID	99VND9
Q9VND9	PRELIMINARY;
PRT;	332 AA.
Q9VND9;	
AC	Q9VND9;
TF	01-MAY-2000 (TREMBLrel. 13, Created)
JT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
VT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
WT	CRMP PROTEIN.
EN	CRMP OR BCDNA;HL02693 OR CG1411.
N	Drosophila melanogaster (Fruit fly).
S	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
C	Prterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
C	Ephydroidea; Drosophilidae; Drosophila.
X	NCBI Taxid=7227;

[1] *Journal of the American Medical Association*, 2019;321(10):1000-1001.

SEQUENCE FROM N. A.

STRAIN-BERKELEY;

MEDLINE-201906006; PubMed=107311132;

Adams M.D., Celnikler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
A Ananaitides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
A George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
A Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
A Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
A Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
A Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
A Ballieu R.M., Basu A., Bayevdale J., Bayraktaroglu L., Beasley E.M.,
A Beeson K.Y., Benos P.V., Bereman B.P., Bhandari P., Bolshakov S.,
A Burks K.C., Butch M.N., Bock J., Brockstein P., Brotter P.,
A Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
A Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.

-Tue Jul 31 13:08:21 2001

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.Y., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltsenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 DR EMBL; AE003602; AAF52003.2;
 DR FlyBase; FBgn0023023; CRM.
 DR InterPro; IPR002195;
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 332 AA; 36453 MW; 97BF4D2CFF3F601 CRC64;

Query Match 28.08; Score 834.5; DB 5; Length 332;
 Best Local Similarity 51.48; Pred. No. 1.9e-52;
 Matches 166; Conservative 50; Mismatches 100; Indels 7; Gaps 3;

QY 7 KSPV---RTSRLLRGRVNDQSFYADVHVHEDGLIKQIG--ENLIVPGGIHTIDAH 61
 DB 9 KVPVHLQSAQNRVYKNGEIVNHDKSKADYVIEDGLIKFVGSSEITIPGGVRTIDAS 68
 QY 62 GLMVLPGGVVHTRQLMPVLGTPADDFCQGTAAALAGTTMILDHVPDPTGVSLAAEY 121
 DB 69 GLMIIPGGIDPHTHQLPFGGAVAVDDFYHGTTAAVAGGTTMIDFVLPNKHESWIEAYD 128
 QY 122 QWRERADSAACDYSLVHVDITRWHSIKIELEALVKRGVNSFLVFMAYKRCQCSQSM 181
 DB 129 KRSWADPKVCCDYGHLVGTWWSKSVSEIGILCKELGVNSFKTFMAYKGLYLQUNSD 188
 QY 182 YEIFSIRDLGALAQAQVHAENGDIIVEEQKRLLELGTITGPEGHVLSHPPEVEAEVYRAV 241
 DB 189 LDVFERIRHLGVAVYHAENGDIIVAKTORLLAEGINGPEGHLSRPERVEAEVYRAV 248
 QY 242 IAKQANCPYVTKVMSKGAADATAQAKR--RGVVFSGEPITASLGTGDSHYWSKNWAKA 299
 DB 249 LAHQADCPYVTVVHMSKAGIELARAHRYRGYINGETLAAALGTDATCCQHLGFDAA 308
 QY 300 AFTVSPVPNDPTTADHLTCLLS 322
 DB 309 AHVLSPLRLPKTTPPEFLMKLLA 331

RESULT 12
 ID O69809 PRELIMINARY; PRT; 467 AA.
 AC O69809;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE D-HYDANTOINASE (EC 3.5.2.2) (DIHYDROPYRIMIDINASE) (DHPASE).
 GN SCIA6.04.
 OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
 CC UREIDOPROPIONATE.
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 DR EMBL; AL023496; CAA18902.1;
 DR InterPro; IPR002195;
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Hydrolyase.
 SQ SEQUENCE 467 AA; 50748 MW; CA4720374DEB0F0B CRC64;

Query Match 23.28; Score 692.5; DB 2; Length 467;
 Best Local Similarity 33.88; Pred. No. 5.9e-42;
 Matches 153; Conservative 84; Mismatches 192; Indels 43; Gaps 10;

QY 14 SDRLLRGRVNDQSFYADVHVHEDGLIKOI-----ENLIVPGGIHTIDAH 61
 DB 2 SSRTVIRGLVITASDEIHADVLIEDGRVAALATGTTPAAEAFTAEV-----IDAS 53
 QY 62 GLMVLPGGVVHTRQLMPVLGTPADDFCQGTAAALAGTTMILDHVPDPTGVSLAAEY 121
 DB 54 GKTVIFGGVDGHTHMEMPPGGTYAADTFETGTRAAAWGGTTTIVDFAIQSVGHSRLREG 113
 QY 122 QWRERADSAACDYSLVHVDITRWHSIKIELEALVKRGVNSFLVFMAYKRCQCSQSM 181
 DB 114 AWHAKAEGNCAIDYGFHMTVSDVNOETLKEMDLLV--EEGVTSFKQFMAYPGVYSDGQI 172
 QY 182 YEIFSIRDLGALAQAQVHAENGDIIVEEQKRLLELGTITGPEGHVLSHPPEVEAEVYRAV 241
 DB 173 LFAMQRAENGGLIMHMAENGIAIDVLVEQALARGETDPRFHGEVRKALLEAEATHRAIR 232
 QY 242 IAKQANCPYVTKVMSKGAADATAQAKRGGVVVFG--PITASLGTGDSHYWSKNWA--- 296
 DB 233 LAQVAGAPLVVHVVSATEAVAELTRADEGLPVFGTGCQYLFLESTD-----NLAE 285
 QY 297 -KAAAFVTSPPVNDPTTADHLTCL---LSSGDLQVTSAGHCTFT--TAQAKVGNKFNALI 351
 DB 286 FEGAKYVCSPTLRP-----KEHQALNRLGLTNDLVVSTDKPCFCFGKELGRGDSRI 341
 QY 352 PEGTNGIEERMSVWKECVASGKMDENEFVAVTSTNAKLFNFYPRKGRVAVGSDADLVI 411
 DB 342 FNGMPGVENRMDLL--HQAVVEGHIGRRRWIEIACATPFMFGLYPKGTITACADADIV 400
 QY 412 WNPATKTIISAKTNLNVYNIPEGVCECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPR 471
 DB 401 YDPHAEQVISAETHMNVDSIAYEGRRITGRVETVLSRGEPPVVTREYTRGKGHGAYTPR 460
 QY 472 KT 473
 DB 461 AT 462

RESULT 15

us-09-367-496-8.std.rspt

•Tue Jul 31 13:08:21 2001

Q9KAH8 PRELIMINARY; PRT: 438 AA.

AC Q9KAH8; 2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

DE ALLANTOININASE.

GN BH2309.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RA Takami H., Nakasone K., Takaki Y.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

EMBL: AP001515; BAB06028.1; -

DR InterPro: IPR002195; -

DR Pfam: PF00744; Dihydroorotase; 1.

DR PROSITE: PS00482; DIHYDROOROTASE.1; UNKNOWN 1.

SQ SEQUENCE 438 AA; 48542 MW; 47A24DD790AF947F CRC64;

Query Match 10.6%; Score 317; DB 2; Length 438;

Best Local Similarity 24.8%; Pred. No. 7.4e-15;

Matches 117; Conservative 80; Mismatches 204; Indels 70; Gaps 16;

Qy 17 LLIRGGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHT-----IDAHGLMVLPGGVD 71

Db 6 LIIRSTVVTETTYRADVAIRGIVSAITE-----PGSISSDDGPAIDGTGLHLFFPGVD 61

Qy 72 VTRLQMPVLGMPADDFCOCTKAALAGTMTILDHVFDPDGVSL- AAYEOWRERADSA 130

Db 62 VHVHNEP--GRTEWEGFASGSKSLAAGVTTYFDMPLNSNPPTITREELDKKQLANEK 119

Qy 131 ACCDYSLVHVDITRWHESTKELEAL--VKEGVNSFLVFMAYKDRCCSDSQMYEIFSII 188

Db 120 SLVDYRF-----WGGLVPGNIDHLQDLHDGVIIGFKAFMS---ECGTDDDFQSHDETLL 170

Qy 189 RD-----LGALAQVHAENGDIIVE-----EOKRLELGITGPEGHVLSHPEEVEAEA 235

Db 171 KGMKIAALGSLAVHAESNMVNALTTIAEEQRLTV-----KDYSEARPIVSELEA 223

Qy 236 VTRAVTIAQANCPLYVTVMKGAADAIAQAKRGV-----VVFGEPTASLGT 285

Db 224 VERILRFAQLTCCPIHICHVSRKVLKRIKQAKGEGVNVSVETCPHYLLSLDEFAEIG- 282

Qy 286 DGSYVWYKKNWAKAAAEVTPPVNPDFTADHLTCLSSGDLQVTGSAHCTTTAQKAVGK 345

Db 283 -----YLAK-----CAPPLRERQVED-LWDGLMAGEIDLISSDHSP-SLPOMKTGK 327

Qy 346 DNFAIPEGTNGTEERMSVMWEKCVASGKMDENEFVATSTNAKIFNFYPRKRVAVGS 405

Db 328 TIFE-VWGGIAGCONTLAVMLTEGYHKRMPLTQIVQLLSTEPAKRFRGLYPQKGTIQVGA 386

Qy 406 DADLVINWPKATKIISAKTNHNLNVEYNIEGVECRGAPAVISQGRVALED 456

Db 387 EASTLLDLNESYTLNASDLRYRHPISPYVQGRFRGKVKHTTCQGRKHYQD 437

Search completed: July 30, 2001, 11:42:00

Job time: 229 sec

Tue Jul 31 13:08:20 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:40:01 ; Search time 10.27 Seconds

(without alignments)
1907.900 Million cell updates/sec

Title: US-09-367-496-8

Perfect score: 2985

Sequence: 1 MSFQKKSIPTSDRLIR.....RTAKINAPPGRSNITSLS 572

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2972	99.6	572	1 DPV4_HUMAN	O14531 homo sapien
2	2818	94.4	572	1 DPV4_MOUSE	O35098 mus musculus
3	2789	93.4	564	1 DPV4_RAT	O62951 rattus norv
4	2362	79.1	572	1 DPV2_CHICK	O80635 gallus gall
5	2347	78.6	572	1 DPV2_BOVIN	O02673 bos taurus
6	2343	78.5	572	1 DPV2_RAT	P47942 rattus norv
7	2339	78.4	572	1 DPV2_HUMAN	O16555 homo sapien
8	2325	77.9	572	1 DPV2_MOUSE	O08553 mus musculus
9	2187	73.3	570	1 DPV3_HUMAN	O14195 homo sapien
10	2172.5	72.8	571	1 DPV3_XENLA	O13022 xenopus lae
11	2172.5	72.8	571	1 DPV3_MOUSE	P97427 mus musculus
12	2157	72.3	572	1 DPV1_MOUSE	O62950 rattus norv
13	2151	72.1	572	1 DPV1_HUMAN	O14194 homo sapien
14	2151	72.1	572	1 DPV1_HUMAN	O14117 homo sapien
15	1554	52.1	519	1 DPVS_HUMAN	O63150 rattus norv
16	1548	51.9	519	1 DPVS_RAT	O62952 rattus norv
17	1451	48.6	572	1 DPV3_RAT	O18677 caenorhabd
18	1245	41.7	520	1 DPV1_CAEL	O21773 caenorhabd
19	1122	37.6	544	1 DPV2_CAEL	O59699 pseudomonas
20	1040	34.8	495	1 HYDA_PSEPU	O45515 bacillus st
21	837	28.0	471	1 HYDA_BACST	O01630 caenorhabd
22	781.5	26.2	854	1 UN33_CAEL	O41884 agrobacteri
23	691.5	23.2	457	1 YGEZ_AGRD	O46806 escherichia
24	672	22.5	465	1 YGEZ_ECOLI	P81006 arthrobacte
25	406.5	13.6	458	1 HYDL_ARTAU	O9rv76 deinococcus
26	325.5	10.9	448	1 ALN_DEIRA	P77671 escherichia
27	311.5	10.4	453	1 ALN_ECOLI	P71809 mycobacteri
28	296.5	9.9	430	1 PYRC_MYCTU	O32137 bacillus su
29	295.5	9.9	446	1 ALN_BACSU	O27199 methanobact
30	295.5	9.9	454	1 PYRC_METH	O9rkus streptomyce
31	257.5	8.6	445	1 ALN_STRCO	O66980 aquifex aeo
32	241.5	8.1	422	1 PYRC_AQUAE	O91437 squalus aca
33	229.5	7.7	2242	1 PYR1_SQUAC	

34	221.5	7.4	423	1	PVRC_METJA	Q58885 methanococ
35	220	7.4	427	1	PVRC_BACCL	P46538 bacillus ca
36	212.5	7.1	2225	1	PYR1_HUMAN	P27708 homo sapien
37	212	7.1	460	1	ALN_YEAST	P32375 saccharomyc
38	206.5	6.9	404	1	PVRC_PYRAB	Q9uxv6 pyrococcus
39	205	6.9	417	1	PVRC_PYRHO	O57740 pyrococcus
40	204.5	6.9	424	1	PVRC_PSEPU	Q59712 pseudomonas
41	204	6.8	427	1	PVRC_LACLE	P48795 lactobacilli
42	204	6.8	2185	1	PYR1_DICDI	P20054 dictyosteli
43	203	6.8	2225	1	PYR1_MESAU	P08955 mesocricetu
44	199.5	6.7	423	1	PVRC_PSEAE	Q51551 pseudomonas
45	197.5	6.6	416	1	PVRC_DEIRA	Q9rv76 deinococcus

ALIGNMENTS

RESULT	1
DPV4_HUMAN	
ID	DPV4_HUMAN STANDARD; PRT; 572 AA.
AC	O14531; O00240;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN) .
GN	DPVSL4 OR ULIP4.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RA	HamaJina N., Kato Y., Kouwaki M., Wada Y., Sasaki M., Nonaka M.;
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE OF 1-553 FROM N.A.
RC	TISSUE=Retina;
RX	MEDLINE=98314496; PubMed=9652388;
RA	Byk T., Ozon S., Sobel A.;
RT	"The ulip family phosphoproteins -- common and specific properties.";
RL	Eur. J. Biochem. 254:14-24(1998).
CC	-I- SURCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC	-I- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; AB006713; BAA21886.1; -
DR	EMBL; Y10976; CAA71872.1; -
DR	HSSP; P18316; 1KRC.
DR	InterPro; IPR002195; -
DR	Pfam; PF00744; Dihydroorotase; 1.
FT	CONFLICT 122 122 R -> O (IN REF. 2).
SEQUENCE	572 AA; 61905 MW; 3E72A33E3BED5BE9 CRC64;

Query Match	99.6%;	Score 2972;	DB 1;	Length 572;
Best Local Similarity	99.7%;	Pred. No. 2.le-207;		
Matches 570;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MSFQKKSIPTSDRLIRGRVNDQSFYADVHVHEDGLIKOIGENLIIVPGGIHTIDA	60	
Db	1	NSFQKKSIPTSDRLIRGRVNDQSFYADVHVHEDGLIKOIGENLIIVPGGIHTIDA	60	
QY	61	HGLMVLPGGVVHTRLOMPVLGWTTPADDFCQGTKAALAGGTTMILDRHVPDPTGVSLAA	120	
Db	61	HGLMVLPGGVVHTRLOMPVLGWTTPADDFCQGTKAALAGGTTMILDRHVPDPTGVSLAA	120	


```

QY 121 EQRERADSAACCDYSLHVDITRWHSIKKEEALVKEGVNSFLVMAVKDRCQSDSQ 180
Db 121 EQRERADSAACCDYSLHVDITRWHSIKKEEALVKEGVNSFLVMAVKDRCQSDSQ 180
QY 181 MYEIFSIRDLGALAOVHAENGDIVVEEOKRLELIGITGEGHVLSPHEVEAEAVYRAV 240
Db 181 MYEIFSIRDLGALAOVHAENGDIVVEEOKRLELIGITGEGHVLSPHEVEAEAVYRAV 240
QY 241 TIAQANCPLYTVTKVMSKGAADAIQAQRGVVVFGEPTTASLGTDGSHYKSNWAKAAA 300
Db 241 TIAQANCPLYTVTKVMSKGAADAIQAQRGVVVFGEPTTASLGTDGSHYKSNWAKAAA 300
QY 301 FVTSPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAAQAVGKDNFALIPGEGNGIEE 360
Db 301 FVTSPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAAQAVGKDNFALIPGEGNGIEE 360
QY 361 RMSVWEKCVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
Db 361 RMSVWEKCVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
QY 421 SAKTHNLNVEYNIFEGVBCRGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTFPDEFYK 480
Db 421 SAKTHNLNVEYNIFEGVBCRGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTFPDEFYK 480
QY 481 RIKARNRLAEIHGVPGRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLNHQSGFS 540
Db 481 RIKARNRLAEIHGVPGRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLNHQSGFS 540
QY 541 LSGSQADDDHIARRTAQKIMAPPGGRSNTISLS 572
Db 541 LSGSQADDDHIARRTAQKIMAPPGGRSNTISLS 572

```

RESULT 2

```

DPY4_MOUSE STANDARD; PRT; 572 AA.
AC 035098; 008886;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYSL4 OR ULIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem 254:14-24(1998).
CC 1- SURCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE DIHYDROXYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB006715; BAA21888.1;
CC MGD; MGI:1349764; Dpysl4.
CC
DR
DR
DR

```

```

DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 125 126 ER -> DG (IN REF. 2).
FT CONFLICT 354 354 G -> V (IN REF. 2).
FT CONFLICT 420 420 F -> I (IN REF. 2).
SQ SEQUENCE 572 AA; 61961 MW; 37671129FC02C7AF CRC64;

```

Query Match

```

Best Local Similarity 94.4%; Score 2818; DB 1; Length 572;
Matches 532; Conservative 21; Mismatches 19; Indels 0; Gaps 0;
QY 1 MSFGKKSIPRITSDRLIRGGRIVNDQSFYADVHVEDGLIKQIGNLIVPGGIHTIDA 60
Db 1 MSFGKKSIPRITSDRLIRGGRIVNDQSFYADVHVEDGLIKQIGNLIVPGGIHTIDA 60
QY 61 HGLMVLPGGVVDVTRLOMPVGLMTDPADDFCQGTAKALAGGTTMILDHVPDTCVSLAAY 120
Db 61 HGLMVLPGGVVDVTRLOMPVGLMTDPADDFCQGTAKALAGGTTMILDHVPDTCVSLAAY 120
QY 121 EQRERADSAACCDYSLHVDITRWHSIKKEEALVKEGVNSFLVMAVKDRCQSDSQ 180
Db 121 EQRERADSAACCDYSLHVDITRWHSIKKEEALVKEGVNSFLVMAVKDRCQSDSQ 180
QY 181 MYEIFSIRDLGALAOVHAENGDIVVEEOKRLELIGITGEGHVLSPHEVEAEAVYRAV 240
Db 181 MYEIFSIRDLGALAOVHAENGDIVVEEOKRLELIGITGEGHVLSPHEVEAEAVYRAV 240
QY 241 TIAQANCPLYTVTKVMSKGAADAIQAQRGVVVFGEPTTASLGTDGSHYKSNWAKAAA 300
Db 241 TIAQANCPLYTVTKVMSKGAADAIQAQRGVVVFGEPTTASLGTDGSHYKSNWAKAAA 300
QY 301 FVTSPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAAQAVGKDNFALIPGEGNGIEE 360
Db 301 FVTSPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAAQAVGKDNFALIPGEGNGIEE 360
QY 361 RMSVWEKCVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
Db 361 RMSVWEKCVASGKMDENEFVAVTSTNAKIFNFPYPRGRVAVGSDADLVINWPKATKII 420
QY 421 SAKTHNLNVEYNIFEGVBCRGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTFPDEFYK 480
Db 421 SAKTHNLNVEYNIFEGVBCRGAPAVVISQGRVALEDGKMEVTPGAGRFVPRKTFPDEFYK 480
QY 481 RIKARNRLAEIHGVPGRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLNHQSGFS 540
Db 481 RIKARNRLAEIHGVPGRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLNHQSGFS 540
QY 541 LSGSQADDDHIARRTAQKIMAPPGGRSNTISLS 572
Db 541 LSGSQADDDHIARRTAQKIMAPPGGRSNTISLS 572

```

RESULT 3

```

DPY4_RAT STANDARD; PRT; 564 AA.
AC 062951;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 3) (CRMP-3) (FRAGMENT).
GN DPYSL4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the

```

Tue Jul 31 13:08:20 2001

```
RT nervous system.
RL J. Neurosci. 16:6197-6207(1996).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED TRANSIENTLY IN DEVELOPING SPINAL
CC CORD AND SELECTIVELY IN THE POSTNATAL CEREBELLUM.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U52103; AAB03281.1; -.
CC HSP; P18316; IIRC.
CC InterPro: IPR002195; -.
CC Pfam: PF00744; Dihydroorotase; 1.
CC NON_TER .1
CC SEQUENCE 564 AA; 61085 MW; 42050891CC1436D2 CRC64;
CC -----
CC Query Match 93.4%; Score 2789; DB 1; Length 564;
CC Best Local Similarity 93.3%; Pred. No. 3.5e-194; Indels 0; Gaps 0;
CC Matches 526; Conservative 21; Mismatches 17;
QY 9 IPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPG 68
DB 1 IPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPG 60
QY 69 GVDVHTRQLMPVLGMPADDFCQGTAAALAGGTTMILDHVPDPGCVSLLAAEYQWRERAD 128
DB 61 GVDVHTRQLMPVLGMPADDFCQGTAAALAGGTTMILDHVPDPGCVSLLAAEYQWRERAD 120
QY 129 SAACCDYSLHVDITWHESIKEELEALVKEGVNSFLVFMAYKDRQCDSQSMVEIFSII 188
DB 121 SAACCDYSLHVDITWHESIKEELEALVKEGVNSFLVFMAYKDRQCDSQSMVEIFSII 180
QY 189 RDLGALVAQVHAENGDIIVEEQKRLLELGITGEGHVLSPHEEVEAEVYRAVTTAKQANC 248
DB 181 RDLGALVAQVHAENGDIIVEEQKRLLELGITGEGHVLSPHEEVEAEVYRAVTTAKQANC 240
QY 249 PLIYTVKMSKGAADALIAQAKRGVVVFGPEITASLGTDGSHVSKNKAFAAFTSPVNV 308
DB 241 PLIYTVKMSKGAADALIAQAKRGVVVFGPEITASLGTDGSHVSKNKAFAAFTSPVNV 300
QY 309 PDPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTINGIERMSVMVWEK 368
DB 301 PDPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTINGIERMSVMVWEK 360
QY 369 CVASGRKMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKTHNLN 428
DB 361 CVASGRKMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKTHNLN 420
QY 429 VEYNIIEGVECGCAPAVVISQGRVALEDCKMFTVPCAGRFVPRKTFPDPFVYKRIKARNL 488
DB 421 VEYNIIEGVECGCAPAVVISQGRVALEDCKMFTVPCAGRFVPRKTFPDPFVYKRIKARNL 480
QY 489 AEIHGVPRLGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFSLSGSQADD 548
DB 481 AEIHGVPRLGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFSLSGSQADD 540
QY 549 HIARRTAQKIMAPPGGRSNTISLS 572
DB 541 HIARRTAQKIMAPPGGRSNTISLS 564
RESULT 4
DPY2_CHICK STANDARD; PRT; 572 AA.
AC Q90635;
DT 15-JUL-1998 (Rel. 36, Created)
```

```
15-JUL-1998 (Rel. 36, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSIN RESPONSE
DE MEDIATOR PROTEIN CRMP-62).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Dorsal root ganglion;
RX MEDLINE=95364923; PubMed=7637782;
RA Coshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
RT "Collapsin-induced growth cone collapse mediated by an intracellular
RT protein related to UNC-33."
RL Nature 376:509-514(1995).
CC -1- FUNCTION: INVOLVED IN NEURONAL GROWTH CONE COLLAPSE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U17277; AAA93200.1; -.
CC HSP; P18316; IIRC.
CC InterPro: IPR002195; -.
CC Pfam: P700744; Dihydroorotase; 1.
CC SEQUENCE 572 AA; 62330 MW; 85DB9E3DD5E54D8D CRC64;
CC -----
QY 1 MSFGGKSIPIRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
DB 1 MSYGGKNIPIRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQIGENLIVPGGVKTKIEA 60
QY 61 HGLMVLPGGVVDHTRQLMPVLGMPADDFCQGTAAALAGGTTMILDHVPDPGCVSLLAA 120
DB 61 HGRVLPGLGIDVHTRFQMPGEGMTSADDFCQGTAAALAGGTTMIDHVPDPGCVSLLTAF 120
QY 121 ECWRERADSAACCDYSLHVDITWHESIKEELEALVKEGVNSFLVFMAYKDRQCDSQ 180
DB 121 DCWRERADSKCCDYSLHVDITWHESIKEELEALVKEGVNSFLVFMAYKDRQCDSQ 180
QY 181 MVEIFSIIRDLGALVAQVHAENGDIIVEEQKRLLELGITGEGHVLSPHEEVEAEVYRAV 240
DB 181 IYEVLSVIRDIGATAQVHAENGDIIVEEQKRLLELGITGEGHVLSPHEEVEAEVYRAV 240
QY 241 TIJAQANCPLYTVKMSKGAADALIAQAKRGVVVFGPEITASLGTDGSHVSKNKAFAA 300
DB 241 TIJANOTNCPLYTVKMSKSAAEVIAQAKRGVVVFGPEITASLGTDGSHVSKNKAFAA 300
QY 301 FVTSPVNPDPPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTINGIEE 360
DB 301 FVTSPVNPDPPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTINGIEE 360
QY 361 RHMVMWEKCVASGRKMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKI 420
DB 361 RUSIWDKAVVTGKMDENOFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKI 420
QY 421 SAKTHNLNVEYNIIEGVECGCAPAVVISQGRVALEDCKMFTVPCAGRFVPRKTFPDPFVYK 480
DB 421 SAKTHNISLEYNIIEGVECGCAPAVVISQGRVALEDCKMFTVPCAGRFVPRKTFPDPFVYK 480
QY 481 RIKARNRLAEIHGVPRLGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRNLHQSGFS 540
```

Db 481 RIKARSLAEIRGVRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHSQGS 540
QY 541 LSGSQADHIIARRTAQKIMAPPGKRSNTISL 571
Db 541 LSGAQIDNIPRRRTQRIIVAPPGGRANITSL 571
RESULT 5
DPY2_BOVIN STANDARD; PRT; 572 AA.
ID DPY2_BOVIN AC 002675;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (NEURAL SPECIFIC PROTEIN
DE NSP60).
GN DPYSL2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamata T.K.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; U83278; AAB80618.1;
DR HSP; P18316; IIRC.
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; 343507ACB9D91BDE CRC64;

Query Match 78.6%; Score 2347; DB 1; Length 572;
Best Local Similarity 75.7%; Pred. No. 3.2e-162;
Matches 432; Conservative 72; Mismatches 67; Indels 0; Gaps 0;
QY 1 MSFOGKKSIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 MSYQKKNIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
QY 61 HGLMVLPGVDVHTLQMPVLCMTADDFOCGTKAALAGGTTMILDHVFPDGTGVSLLAAY 120
Db 61 HSRWVPGGIDVHTLQMPVLCMTADDFOCGTKAALAGGTTMILDHVFPDGTGVSLLAAY 120
QY 121 EQRERADSAACCDYSLHVDITRWHEISKELEALVEKGVNSFLVPMAYKDRQCSDSQ 180
Db 121 DQREWADSKCCDYSLHVDITRWHEISKELEALVEKGVNSFLVPMAYKDRQCSDSQ 180
QY 181 MYEFSITRDGALQAQVAENGDIIVEEQKRLLELGTGPEGHVLSHPVEAEAVYRV 240
Db 181 IYELSVIRDIGAQAQVAENGDIIVEEQKRLLELGTGPEGHVLSHPVEAEAVYRV 240
QY 241 TIAQANCPVTVTKVMSKGADATQAARKGVVVFGPEPTASLGTGSGHYSKWNKAKAAA 300
Db 241 TIANQNCPLVITVTKVMSKRAEVAQAARKKGTGVYGEPTASLGTGSGHYSKWNKAKAAA 300
QY 301 FVTSPPVNPDPPTADHLTCLLSGDLQVTSACHTFTTAQKAVGKNEALIPGTNGIEE 360
Db 301 FVTSPPVNPDPPTADHLTCLLSGDLQVTSACHTFTTAQKAVGKNEALIPGTNGIEE 360

QY 361 RMSWWEKCVASKMDENEFVAVTSTNAKIFNFPKGRVAVGSDADLVINPKATKII 420
Db 361 RMSVWIDKAVVTGKMDENQFVAVTSTNAKVENLYPRKRIAGVSDADLVINPDPSVKTI 420
QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISOGRVALEDEGKMFVTPGAGRFVPRKTFDFEYK 480
Db 421 SAKTHNSSELYNIFEGVECRGAPAVVISOGRVILEDEGTLHVTGSGRVTIPKPFDFEYK 480
QY 481 RIKARNRLAEIHGVRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHSQGS 540
Db 481 RIKARSLAEIRGVRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHSQGS 540
QY 541 LSGSQADHIIARRTAQKIMAPPGKRSNTISL 571
Db 541 LSGAQIDNIPRRRTQRIIVAPPGGRANITSL 571
RESULT 6
DPY2_RAT STANDARD; PRT; 572 AA.
ID DPY2_RAT AC P47942;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (TURNED ON AFTER
DE DIVISION, 64 KDA PROTEIN) (TOAD-64) (COLLAPSE IN RESPONSE MEDIATOR
DE PROTEIN 2) (CRMP-2).
GN DPYSL2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 136-142; 402-418; 441-450 & 499-511.
RA STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96033765; PubMed=7472434;
RA Minturn J.E., Fryer H.J.L., Geschwind D.H., Hockfield S.;
RT "TOAD-64, a gene expressed early in neuronal differentiation in the
RT rat, is related to unc-33, a C. elegans gene involved in axon
RL J. Neurosci. 15:6757-6766(1995).
CC -1- FUNCTION: MAY HAVE A ROLE IN AXON ELABORATION.
CC -1- SUBCELLULAR LOCATION: TIGHTLY, BUT NONCOVALENTLY, ASSOCIATED WITH
CC MEMBRANES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY AFTER NEURONAL BIRTH
CC AND IS DRAMATICALLY DOWNREGULATED IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; Z46882; CAA86981.1;
DR InterPro; IPR002195;
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; C031F3BC038AA737 CRC64;

Query Match 78.5%; Score 2343; DB 1; Length 572;
Best Local Similarity 75.8%; Pred. No. 6.2e-162;
Matches 433; Conservative 70; Mismatches 68; Indels 0; Gaps 0;
QY 1 MSFOGKKSIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 MSYQKKNIPRITSRLIRGRVNDQSFYADVHVHEDGLIKQIGENLIVPGGIHTIDA 60
QY 61 HGLMVLPGVDVHTLQMPVLCMTADDFOCGTKAALAGGTTMILDHVFPDGTGVSLLAAY 120
Db 61 HSRWVPGGIDVHTLQMPVLCMTADDFOCGTKAALAGGTTMILDHVFPDGTGVSLLAAY 120

RT RT
DNA Res. 6:291-297(1999).
CC - SURCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC - TISSUE SPECIFICITY: UBIQUITOUS.
CC - SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U17279; AAA93202.1; -
CC EMBL: D78013; BAA1191.1; -
CC EMBL: U97105; AAC05793.1; -
CC EMBL: AH020777; BAA86991.1; JOINED.
CC EMBL: AH020764; BAA86991.1; JOINED.
CC EMBL: AH020765; BAA86991.1; JOINED.
CC EMBL: AH020766; BAA86991.1; JOINED.
CC EMBL: AH020767; BAA86991.1; JOINED.
CC EMBL: AH020768; BAA86991.1; JOINED.
CC EMBL: AH020769; BAA86991.1; JOINED.
CC EMBL: AH020770; BAA86991.1; JOINED.
CC EMBL: AH020771; BAA86991.1; JOINED.
CC EMBL: AH020772; BAA86991.1; JOINED.
CC EMBL: AH020773; BAA86991.1; JOINED.
CC EMBL: AH020774; BAA86991.1; JOINED.
CC EMBL: AH020775; BAA86991.1; JOINED.
CC EMBL: AH020776; BAA86991.1; JOINED.
CC HSP; P18316; IIRC.
CC MIN; 602463; -
CC InterPro; IPR002195; -
CC Pfam; PF00744; Dihydroorotase; 1.
CC
CC SEQUENCE 572 AA; 62293 MW; 5CDB6CF7F5C308AD CRC64;
Query Match 78.4% Score 2339; DB 1; Length 572;
Best Local Similarity 75.5% Pred. No. 1.2e-161;
Matches 431; Conservative 72; Mismatches 68; Indels 0; Gaps 0;
Qy 1 MSFGKKSIPRITSDRLIRGRVINDQSFYADVHVEDGLIKQIGENLVPGGIHTIDA 60
Db 1 MSYQKKNIPRITSDRLIKGKIVNDQSFYADIVMEDGLIKQIGENLVPGGVKTLIA 60
Qy 61 HCLMVLPGVDVHTRLQPLVGMTPADDFCOGTAALAGGTTMLDHFVDPDTGVSLLAA 120
Db 61 HSRMVPVGGIDVHTRFQMPDQGMTSADDFFOGTAALAGGTTMLDHFVDPDTGVSLLAA 120
Qy 121 EQWRERADSAACDYSILHVDITRWHESIKEEALVKEGVNSFLVFMAYKDRCCQSDSQ 180
Db 121 DQWREWADSKSCDYSILHVDISEWHKIQEEMALVKDHGVNSFLVFMAYKDRFQLTDCQ 180
Qy 181 MYEIFSIRDLGALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRA 240
Db 181 IYEVLSVIRDIGALAQVHAENGDIITAEEOQRIIDLGITGPEGHVLSRPEVEAEAVNRAI 240
Qy 241 TIAQANCPLYTVKVMKGAADAIQAQKRRGVVFGPEPITASLGTDGSHYKSNWAKAAA 300
Db 241 TIANQNCPLIYTVKVMKSSAEVIAQARKKGVVYGEPIITASLGTDGSHYKSNWAKAAA 300
Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360
Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360
Qy 361 RSMNVKCVASCKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
Db 361 RSMNVKCVASCKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
Qy 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK 480
Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK 480

Qy 121 EQWRERADSAACDYSILHVDITRWHESIKEEALVKEGVNSFLVFMAYKDRCCQSDSQ 180
Db 121 DQWREWADSKSCDYSILHVDITRWHESIKEEALVKEGVNSFLVFMAYKDRFQLTDCQ 180
Qy 181 MYEIFSIRDLGALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRA 240
Db 181 IYEVLSVIRDIGALAQVHAENGDIITAEEOQRIIDLGITGPEGHVLSRPEVEAEAVNRAI 240
Qy 241 TIAQANCPLYTVKVMKGAADAIQAQKRRGVVFGPEPITASLGTDGSHYKSNWAKAAA 300
Db 241 TIANQNCPLIYTVKVMKSSAEVIAQARKKGVVYGEPIITASLGTDGSHYKSNWAKAAA 300
Qy 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360
Db 301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSACHTFTTAQKAVGKDNFALIPGTINGIEE 360
Qy 361 RSMNVKCVASCKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
Db 361 RSMNVKCVASCKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
Qy 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK 480
Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK 480
Qy 481 RIKARNLAEHTGVRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRNLHQSGFS 540
Db 481 RIKARNLAEHTGVRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRNLHQSGFS 540
Qy 541 LSGSQADHIAARRTAQKIMAPPGGRSNTSL 571
Db 541 LSGAQIDDNIPRRTTORIVAPPGGRANITSL 571
RESULT 7
DPY2_HUMAN STANDARD; PRT; 572 AA.
AC Q16555; O00424;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSEIN RESPONSE
DE MEDIATOR PROTEIN 2) (CRMP-2) (N2A3).
OS Homo sapiens (Human).
GN Dpysl2 Or CRMP2.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95364923; PubMed=7637782;
RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
RT "Collapsin-induced growth cone collapse mediated by an intracellular
RT protein related to UNC-33";
RL Nature 376:509-514(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RT related proteins with differential tissue distribution";
RL Gene 180:157-163(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Liver;
RA Zhou J., Chen Y., Gu J.R.;
RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20039612; PubMed=10574455;
RA Kitamura K., Takayama M., Hamajima N., Nakanishi M., Sasaki M.,
RA Endo Y., Takemoto T., Kimura H., Iwaki M., Nonaka M.

```

Db 421 SAKTHNSLEYNIFEGMECRGSPVIVISQKIVLEDTGLHVTGSGRYIPRKPPDFVYK 480
QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
Db 481 RIKARSRLAEHLRGVPRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHQSGFS 540
QY 541 LSGSQADHDHARRTAOKIMAPPGRSNTLSL 571
Db 541 LSGAQIDDNIPRTTQRIVAPPGRANITSL 571

RESULT 8
DPY2_MOUSE
ID DPY2_MOUSE STANDARD; PRT; 572 AA.
AC O08553;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE DHYDROXYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (ULIP 2 PROTEIN).
GN DPYSL2 OR ULIP2 OR CRMP2.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314456; PubMed=96523388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10339; CAA71370.1; -
DR MGI: 1349763; Dpysl2.
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62170 MW; 7FD4E8A242ACF62D CRC64;

Query Match 77.9%; Score 2325; DB 1; Length 572;
Best Local Similarity 75.3%; Pred. No. 1.2e-160;
Matches 430; Conservative 70; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSFGQKKSIPRTSDRLIRGRIYNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
Db 1 MSYQCKKNIPRTSDRLIRGRIYNDQSFYADYIMEDGLIKOIGENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRQLMPVLGMPADDFCQGTAAALAGGTTMLDHFVDPDTGVSLLAA 120
Db 61 HSRMVIPGGIDVHTRFQMPDQGMTSADDFCQGTAAALAGGTTMLDHFVDPDTGVSLLAA 120
QY 121 EQRERADSAACCDYSLHVDITRWHESIKEELALVKEGVNSFLVPMAYKDCQCSDSQ 180
Db 121 DQREWADSKCCDYSLHVDITRWHESIKEELALVKEGVNSFLVPMAYKDCQCSDSQ 180
QY 181 MYEFTSIIRDLAGAQAQVHAENGDIIVEEQRKLLLELGTITGPEGVHLSHPERVEAEVYRAV 240
Db 181 IYEVLSVIRDTGALQAQVHAENGDIIEAQAQRIILDGLITGPEGVHLSHPERVEAEVYRAV 240
QY 241 TIAQANCPLYVTKVMSKGAADAIAQAKRGVVVVEPTASLGTDGSHYWSKNWAKAAA 300
Db 241 TIANQTNCPLYVTKVMSKGAADAIAQAKRGVVVVEPTASLGTDGSHYWSKNWAKAAA 300

```

```

QY 301 FVTSPPVPDPPTADHLTCLLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
Db 301 FVTSPPPLSPDPPTDFLNSLLSCGDLQVTSAGHCTENTTAQKAVGKDNFALIPETNGIEE 360
QY 361 RSMWWEKCVASGKMDENEFVAVTSTNAKIFENFYPRKGRVAVGSDADLVINWPKATKII 420
Db 361 RSMVINDKAVVTGKMDENEFVAVTSTNAKIFENFYPRKGRISVGSDADLVINWPKATKII 420
QY 421 SAKTHNLVNEYIFEGVECGAPVAVVISOGRVVALEDGKMFVTPGAGREVPRTKTEPDVYK 480
Db 421 SAKTHNSALEYIFEGMECRGSPVIVISQKIVLEDTGLHVTGSGRYIPRKPPDFVYK 480
QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSGFS 540
Db 481 RIKARSRLAEHLRGVPRGLYDGPVCEVSVTPKTVTPASSAKTSPAKQAPPVRLNHQSGFS 540
QY 541 LSGSQADHDHARRTAOKIMAPPGRSNTLSL 571
Db 541 LSGAQIDDNIPRTTQRIVAPPGRANITSL 571

```

```

RESULT 9
DPY3_MOUSE
ID DPY3_MOUSE STANDARD; PRT; 570 AA.
AC Q62188;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DHYDROXYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (UNC-33-LIKE
DE PHOSPHOPROTEIN) (ULIP PROTEIN).
GN DPYSL3 OR ULIP OR DRP3.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=96133606; PubMed=8551352;
RA Byk T., Dobransky T., Cifuentes-Diaz C., Sobel A.;
RT "Identification and molecular characterization of Unc-33-like
RT phosphoprotein (Ulip), a putative mammalian homolog of the axonal
RT guidance-associated unc-33 gene product.";
RL J. Neurosci. 16:688-701(1996).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X87817; CAA61082.1; -
DR MGI: 1349762; Dpysl3.
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
SQ SEQUENCE 570 AA; 61936 MW; 14C339BF90015D96 CRC64;

```

```

Query Match 73.3%; Score 2187; DB 1; Length 570;
Best Local Similarity 70.1%; Pred. No. 1.2e-150;
Matches 401; Conservative 90; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSFGQKKSIPRTSDRLIRGRIYNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDA 60
Db 1 MSYQCKKNIPRTSDRLIRGRIYNDQSFYADYIMEDGLIKOIGENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRQLMPVLGMPADDFCQGTAAALAGGTTMLDHFVDPDTGVSLLAA 120
Db 61 NGKXVIPGGIDVHTRFQMPYKGMTTVDFFCQGTAAALAGGTTMLDHFVDPDTGVSLLAA 120

```

CC		use by non-profit institutions as long as its content is in no way	
CC		modified and this statement is not removed. Usage by and for commercial	
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC		or send an email to license@isb-sib.ch)	
CC		-----	
DR	EMBL; D78014; BAAL1192.1; -		
DR	EMBL; Y07818; CAA69153.1; -		
DR	MIM; 601168; -		
DR	InterPro; IPR002195; -		
DR	Pfam; PF00744; Dihydroorotase; 1. L -> V (IN REF. 2).		
FT	CONFLICT 49 49		
FT	CONFLICT 142 142 T -> A (IN REF. 2).		
FT	SEQUENCE 570 AA; 61963 MW; 9D6AF86CB33AD5 CRC64;		
Query Match		73.2%; Score 2186; DB 1; Length 570;	
Best Local Similarity		69.9%; Pred. No. 1.4e-150;	
Matches 400; Conservative 91; Mismatches 79; Indels 2; Gaps 2;			
QY	1 MSQGGKSIPTSDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTDA 60		
DB	1 MS:QGGKSIPTSDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTDA 60		
QY	61 HGLMVLPGGVYVHTRQLMPVLGMPADDFCQGTAAALAGTTMLDHFVDTGVSLAA 120		
DB	61 NGKMWIPGGIDVHTFQMPYKGMTTVDFFQGTAAALAGTTMLDHFVDTGVSLAA 120		
QY	121 EQRERADSAACDYSLHVDITRWHESIKKELEALYKGVNSFLVPMAYKDRCCSDSQ 180		
DB	121 EKWREWADGSCCDYALHVDITRWHESIKKELEALYKGVNSFLVPMAYKDRCCSDSQ 180		
QY	181 MYEFTSIIRDLGALAAQVHAENGDIVEEQRKLELGTGEGHVLSPHEVEAEAVYAV 240		
DB	181 LYEFTCLGELGALAAQVHAENGDIIEAQEQLRLELGTGEGHVLSPHEVEAEAVFRAI 240		
QY	241 TIAQANCPYVTVKMSKAADATAQAKRGVYVVEGPEITASLCTDGSYKSNWAKAAA 300		
DB	241 TIASQTCNCPYVTVKMSKAADLISQARKGNVYVVEGPEITASLCTDGSYKSNWAKAAA 300		
QY	301 FVTSPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360		
DB	301 FVTSPPLSPDPTPDYINSLASGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGVVEE 360		
QY	361 RSMVWEKCVASGKMDNEFVAVSTNAKIFNYPKRGVAVGSDADLVINPKATKII 420		
DB	361 RSMVWIDKAVATGKMDENQFVAVSTNAKIFNLYPKRGVAVGSDADLVINPKATKII 420		
QY	421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTFPDFYK 480		
DB	421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTFPDFYK 480		
QY	481 RIKARNRLAEITHGVPRGLYDGPVHEVMYPAKPGSGAPARASCPGKISVPPVRLNHSQGS 540		
DB	481 RIKARKKADLHVPGRMGVDFVDTTTPKGGTSGAGSPTSPN-PPVRLNHSQGS 539		
QY	541 LSGSQADHIIARTAKIMAPPGSRNITSLS 572		
DB	540 LSGTQVDEGV-RSASKRIVAPPGSRNITSLS 570		
RESULT 11			
DPY3_XENLA			
ID	DPY3_XENLA	STANDARD;	PRT; 571 AA.
AC	OL3022;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (NEURAL SPECIFIC PROTEIN		
DE	1).		
GN	NSPI.		
OC	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		

QY	121 EQRERADSAACDYSLHVDITRWHESIKKELEALYKGVNSFLVPMAYKDRCCSDSQ 180		
DB	121 EKWREWADGSCCDYALHVDITRWHESIKKELEALYKGVNSFLVPMAYKDRCCSDSQ 180		
QY	181 MYEFTSIIRDLGALAAQVHAENGDIVEEQRKLELGTGEGHVLSPHEVEAEAVYAV 240		
DB	181 LYEFTCLGELGALAAQVHAENGDIIEAQEQLRLELGTGEGHVLSPHEVEAEAVFRAI 240		
QY	241 TIAQANCPYVTVKMSKAADATAQAKRGVYVVEGPEITASLCTDGSYKSNWAKAAA 300		
DB	241 TIASQTCNCPYVTVKMSKAADLISQARKGNVYVVEGPEITASLCTDGSYKSNWAKAAA 300		
QY	301 FVTSPVNPDPPTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360		
DB	301 FVTSPPLSPDPTPDYINSLASGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGVVEE 360		
QY	361 RSMVWEKCVASGKMDNEFVAVSTNAKIFNYPKRGVAVGSDADLVINPKATKII 420		
DB	361 RSMVWIDKAVATGKMDENQFVAVSTNAKIFNLYPKRGVAVGSDADLVINPKATKII 420		
QY	421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTFPDFYK 480		
DB	421 SAKTHNLNVEYNIFEGVRCGAPAVVISQGRVALEDGKMFVTPGAGREVPKRTFPDFYK 480		
QY	481 RIKARNRLAEITHGVPRGLYDGPVHEVMYPAKPGSGAPARASCPGKISVPPVRLNHSQGS 540		
DB	481 RIKARKKADLHVPGRMGVDFVDTTTPKGGTSGAGSPTSPN-PPVRLNHSQGS 539		
QY	541 LSGSQADHIIARTAKIMAPPGSRNITSLS 572		
DB	540 LSGTQVDEGV-RSASKRIVAPPGSRNITSLS 570		
RESULT 10			
DPY3_HUMAN			
ID	DPY3_HUMAN	STANDARD;	PRT; 570 AA.
AC	Q14195; Q93012;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (UNC-33-LIKE		
DE	PHOSPHOPROTEIN) (ULIP PROTEIN).		
GN	DRP3 OR ULIP OR DRP3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RC	MEDLINE=97128821; PubMed=8973361;		
RA	Gaetano C., Matsuo T., Thiele C.J.;		
RA	"A novel gene family defined by human dihydropyrimidinase and three		
RT	related proteins with differential tissue distribution."		
RL	Gene 180:157-163(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RC	MEDLINE=9727371; PubMed=9115293;		
RA	Gaetano C., Matsuo T., Thiele C.J.;		
RT	"Identification and characterization of a retinoic acid-regulated		
RT	human homologue of the unc-33-like phosphoprotein gene (hulip) from		
RT	neuroblastoma cells."		
RL	J. Biol. Chem. 272:12195-12201(1997).		
CC	-1- SUBCELLULAR LOCATION: CYTOSOL.		
CC	-1- TISSUE SPECIFICITY: MAINLY IN HEART AND SKELETAL MUSCLE. ALSO		
CC	EXPRESSED STRONGLY IN FETAL BRAIN AND SPINAL CORD.		
CC	-1- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		


```
CC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Engel E.R., Lepperdinger G., Richter K.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y13069; CAA73509.1; -
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
SQ SEQUENCE 571 AA; 62099 MW; 30A841522E27743D CRC64;

Query Match 72.8%; Score 2172.5; DB 1; Length 571;
Best Local Similarity 69.8%; Pred. No. 1.3e-149;
Matches 399; Conservative 85; Mismatches 87; Indels 1; Gaps 1;

QY 1 MSFGKKSIPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQICENLIVPGGIHTIDA 60
DB 1 MSYQKKNIPRITSERLLIKGGIRVNDQSFYADIWEDGLIKQICENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRLOMPVLGMPADDFCQGTAAALAGTTMILDHVFPDGVSLAAY 120
DB 61 NGKMWIPGGIDVHTLOMPYRGTTVDFFLQGTAAALAGTTMIVDHVPEASLTFEAF 120
QY 121 EOWRERADSAACCDYSLHVDITRWHESTKEEALVKEKGVNSFLVFMAYKDRCCQSDSQ 180
DB 121 EKREWADGKTCCDYSLVHDIHWSVSVQEVETLVKQGVNSFMVYMAKDYQMSNTE 180
QY 181 MYEIFSIRDLGALQAQVHAENGDIIVEEQKRLLELGIITGPEGHVLSHPPEEAEAVYRAV 240
DB 181 LYEIFTFLGLGALQAQVHAENGDIIVEEQKRLLELGIITGPEGHVLSHPPEEAEAVYRAV 240
QY 241 TIQAQNCPLYVTVKMSKGAADAIAQAKRGVVFGEPTASLGTGDSHYWKNWAKAAA 300
DB 241 ITASQTCNPLYVTVKMSKSSVDLIISQARKKGVVTFGEPTASLGTGDSHYWKNWAKAAA 300
QY 301 FVTSPPVNPDPPTADHLTCLLSGDLQVTSAGHCTFTTAQKAVGKDNFALIEGTVNGIEE 360
DB 301 FVTSPPVNPDPPTADHLTCLLSGDLQVTSAGHCTFTTAQKAVGKDNFALIEGTVNGIEE 360
QY 361 RMSVWKEKVASCKMDENEFVAVTSTNAKIFNTPYKRGVAVGSDADLVINPKATKII 420
DB 361 RMSVWKEKVASCKMDENEFVAVTSTNAKIFNTPYKRGVAVGSDADLVINPKATKII 420
QY 421 SAKTHNLAVENIFEGVCEKCAPAVVISQGRVLEDEGKMEVTPGAGRFVPRKTPDFPVYK 480
DB 421 SAKSHSAAENYIPGEMELRGAPLVICQKIMMEDGTLHVTQGRTPCSPFPDYVYK 480
QY 481 RIKARNRLAETHGVPRGLYDGPVHEVWYPAKPGSGAPARASCPGKISVPPVRLNHSQGS 540
DB 481 RIKARKMAELHGVPRGMDYDGVLDLASTPRAGTPAGSTKGSPTK-QTPPVRLNHSQGS 540
QY 541 LSGQADDHIAARTAAQKIMAPPGGRSNTITSL 572
DB 540 LAGNOGDESGVRSARRIVAPLGAPNITSL 571

RESULT 12
DPY1_MOUSE
ID DPY1_MOUSE STANDARD; PRT; 572 AA.
AC P97427; O08554; O35097;
DT 15-JUL-1998 (Rel. 36, Created)
```

```
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROXYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSIDIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1) (ULIP3 PROTEIN).
GN CRMP1 OR DPYSL1 OR ULIP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97262103; PubMed=9107681;
RA Cohen-Salmon M., Crozet F., Rebillard G., Petit C.;
RT "Cloning and characterization of the mouse collapsin response
RT mediator protein-1, Crmp1";
RL Mamm. Genome 8:349-351(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROXYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U72875; AAB39703.1; -
DR EMBL: Y09080; CAA70300.1; -
DR EMBL: AB006714; BAA21887.1; -
DR MGI: 107793; Crmp1.
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 476 476 E -> K (IN REF. 3).
FT CONFLICT 489 489 F -> S (IN REF. 3).
FT CONFLICT 520 520 K -> E (IN REF. 2).
SQ SEQUENCE 572 AA; 62167 MW; FE17DDCD735CAF8F CRC64;

Query Match 72.3%; Score 2157; DB 1; Length 572;
Best Local Similarity 69.2%; Pred. No. 1.7e-148;
Matches 395; Conservative 85; Mismatches 91; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSDRLIRGGIRVNDQSFYADVHVEDGLIKQICENLIVPGGIHTIDA 60
DB 1 MSYQKKNIPRITSERLLIKGGIRVNDQSFYADIWEDGLIKQICENLIVPGGVKTIEA 60
QY 61 HGLMVLPGGVVHTRLOMPVLGMPADDFCQGTAAALAGTTMILDHVFPDGVSLAAY 120
DB 61 NGKMWIPGGIDVHTLOMPYRGTTVDFFLQGTAAALAGTTMIVDHVPEASLTFEAF 120
QY 121 EOWRERADSAACCDYSLHVDITRWHESTKEEALVKEKGVNSFLVFMAYKDRCCQSDSQ 180
DB 121 EKREWADGKTCCDYSLVHDIHWSVSVQEVETLVKQGVNSFMVYMAKDYQMSNTE 180
QY 181 MYEIFSIRDLGALQAQVHAENGDIIVEEQKRLLELGIITGPEGHVLSHPPEEAEAVYRAV 240
DB 181 LYEIFTFLGLGALQAQVHAENGDIIVEEQKRLLELGIITGPEGHVLSHPPEEAEAVYRAV 240
QY 241 TIQAQNCPLYVTVKMSKGAADAIAQAKRGVVFGEPTASLGTGDSHYWKNWAKAAA 300
```


Tue Jul 31 13:08:20 2001

Matches 393; Conservative 87; Mismatches 91; Indels 0; Gaps 0;

```

1 MSFOGKKSIPRITSDRLIRGRIVNDQSFYADVHVDEGLIKOIGENLIVPGIHTIDA 60
1 MSHOGKKSIPHTSDRLIRGRIVNDQSFYADVHVDEGLIKOIGENLIVPGVKTIEA 60
61 HGLVLPVGVVHYFLOMPVLMTPADDFCOGTRKAALAGGTTMILDHVFPDPTGVSLAAY 120
61 NGRMVPVIGSDVNTVYLQKPSQGMTSADDFGOTRAALAGGTTMIDHVPVPGSSLTSE 120
121 EOWHERADSAACCDYSLHVDITRWHESESKEELALVKEGVNSFLVMAYKDRCCQSDSQ 180
121 EKWHIEAADTKSCDYSLHVDITRWHESESKEELALVKEGVNSFLVMAYKDRCCQSDSQ 180
181 MYEETPSIIRDLGALAQVHAENGDIVEEQKRLLELGTITGPEGHVLSHPVEEAEAVYRAV 240
181 LYEAFTFLKGLGAVILVHAENGDLIAEQKRLLELGTITGPEGHVLSHPVEEAEAVYRAV 240
241 TIAKQANCPVYITKVMKSAADIIALARKKPLVFEPTAAASLGTDTGTHYWSKNWAKAAA 300
241 AIAGRINCPVYITKVMKSAADIIALARKKPLVFEPTAAASLGTDTGTHYWSKNWAKAAA 300
301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTTTAQKAVGKDNFALIEPENGIEE 360
301 FVTSPLSPDPPTDYLTSLACGDLQVTSAGHCTTTAQKAVGKDNFALIEPENGIEE 360
361 RMSVMWEKCVASGKMDNEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
361 RMTVVWDKAVATGKMDENQFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
421 SAKTHNLNVEYNIFEGVECGRGAFAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
421 TAKSHKSTVEYNIFEGVECGRGAFAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSAPARASCPGKISVPPVRLNHSQGSF 540
481 RVIIRSKVFLHSVSRGMDYGPVEVPATPKHAAPASAKSSPSKHPHPPPIRLNHSQGSF 540
541 LSGSQADHIIARTAQKIMAPGGRSNTSL 571
541 LSGAQIDNNPRRTGHRIVAPPGGRSNTSL 571

```

RESULT 14
ID DPY1_HUMAN STANDARD; PRT; 572 AA.
AC Q14194; Q13024;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1).
GN CRMP1 CR DPYSL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RL related proteins with differential tissue distribution.";
RL Gene 180:157-163(1996).
RN [2]
RP SEQUENCE OF 64-572 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95364923; PubMed=7637782;
RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
RT "Collapsin-induced growth cone collapse mediated by an intracellular
RL protein related to UNC-33.";
RL Nature 376:509-514(1995).

```

241 ATAGRINCPVYITKVMKSAADIIALARKKPLVFEPTAAASLGTDTGTHYWSKNWAKAAA 300
301 FVTSPPVNPDPPTADHLTCLSSGDLQVTSAGHCTTTAQKAVGKDNFALIEPENGIEE 360
301 FVTSPLSPDPPTDYLTSLACGDLQVTSAGHCTTTAQKAVGKDNFALIEPENGIEE 360
361 RMSVMWEKCVASGKMDNEFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
361 RMTVVWDKAVATGKMDENQFVAVTSTNAAKIFNYPKRGVAVGSDADLVINPRATKII 420
421 SAKTHNLNVEYNIFEGVECGRGAFAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
421 TAKSHKSTVEYNIFEGVECGRGAFAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
481 RIKARNRLAEIHGVRGLYDGPVHEVMVPAKPGSAPARASCPGKISVPPVRLNHSQGSF 540
481 RVIIRSKVFLHSVSRGMDYGPVEVPATPKHAAPASAKSSPSKHPHPPPIRLNHSQGSF 540
541 LSGSQADHIIARTAQKIMAPGGRSNTSL 571
541 LSGAQIDNNPRRTGHRIVAPPGGRSNTSL 571

```

RESULT 13
ID DPY1_RAT STANDARD; PRT; 572 AA.
AC Q62950; P70546;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1).
GN CRMP1 OR DPYSL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
RL nervous system.";
RL J. Neurosci. 16:6197-6207(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Quach T.T., Honnorat J., Aguerre M., Belin M.F., Kolattukudy P.E.,
RA Antoine J.C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U52102; AAB03280.1; -;
DR EMBL; U52095; AAB07042.1; -;
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 3 H -> Y (IN REF. 2).
SQ SEQUENCE 572 AA; 62195 MW; ED63BD8C751CCPDF CRC64;
Query Match 72.1%; Score 2152; DB 1; Length 572;
Best Local Similarity 68.8%; Pred. No. 4e-148;

```

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -2- TISSUE SPECIFICITY: BRAIN.
CC -3- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D78012; BAA11190.1; -;
DR EMBL; U17278; AAA93201.1; -;
DR MIM; 602462; -;
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 504 504 Y -> H (IN REF. 2).
SQ SEQUENCE 572 AA; 62183 MW; A5385FCC79328A30 CRC64;
-----
Query Match 72.1%; Score 2151; DB 1; Length 572;
Best Local Similarity 59.0%; Pred. No. 4.7e-148;
Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;
QY 1 MSFQKKSIPIRISDRLLIRGGRTVNDQSFYADVHVDEGLIKQIGENLIVPGGIHTIDA 60
DB 1 MSYQKKSIPIHTISDRLLIRGGRIINDQSLYADVLEGLIKQIGENLIVPGVKITIEA 60
QY 61 HGLMVLPGGVVDVHTRLOMPVLGTPADDFCQGTAAALAGGTTMLDHFVFDTPGVSLAAY 120
DB 61 NGRMVLPGGIDVNTYLQPKPSOGTAAADDFCQGTAAALVCGTTMLDHFVPEGSSLLTSF 120
QY 121 EOWREDAASACDYSILHVDITRWHESIKEELEALYKEGVNSFLVFMAFKDRCCQSDSQ 180
DB 121 EKWEHAADTKSCDYSILHVDITSWDGVRELEVLVDQKGVNSFQVVMAYKDVYQMSDSQ 180
QY 181 MTEIFSIIRDLAGAQAQVHAENGDIIVEEOKRLELGITGPEGHVLSPHEVEAEVYRAV 240
DB 181 LYEAFTFLKGLGAVILVHAENGDLTAQEQRIEMGITGEGHALSRPELEAEAVFRAI 240
QY 241 TIAQANCLPVYTKVWSKCAADAIAQAKRGVVVFGEPIATASLGTDGSHYWSKNWAKAAA 300
DB 241 ITAGINCPVYITKWSKSAADIIALARKGPLVFGEPIATAASLGTDGTHYWSKNWAKAAA 300
QY 301 FVTSPPVNDPPTADHLTCLSSGDLQVTSQAHCTFTTAQKAVGKDNFALPEGTNGIEE 360
DB 301 FVTSPLSPDPTPDYLTSLACGDLQVTSQCHCPYSTAQKAVGKDNFTLPEGVNGIEE 360
QY 361 RMSVWEKCVASGKMDENEFYAVVSTNAKTFNYPKRGVAVGSDADLVINPKATKII 420
DB 361 RMTVVDKAVATGMDENQFVAVVSTNAKTFNLYPRKGRIAVGSDADVDVLPDKLTI 420
QY 421 SAKTHNLNVEYNIFEVECRGAPAVISQGRVVALEDGKMFVTPCAGRFVPRKTPDFVYK 480
DB 421 TAKSHKSAVEYNIFEVECHSGPLVVISQGIKIVEDGNINVMKGMGRFIRPKAFPEHLYQ 480
QY 481 RIKARNRLAEHTGVPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISGVPVRNLHQSGFS 540
DB 481 RVKLRNVFGLQGVSRGMDYDGPVVEVPATPKYATPAPSAKSSPKHQPPPIRNLHQSNFS 540
QY 541 LSGSQADDHIARTATAQIMAPPGGRSNTSL 571
DB 541 LSGAQDDNNPRTGTHRIVAPPGGRSNTSL 571
-----
RESULT 15
ID DPYS_HUMAN
AC Q14117;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

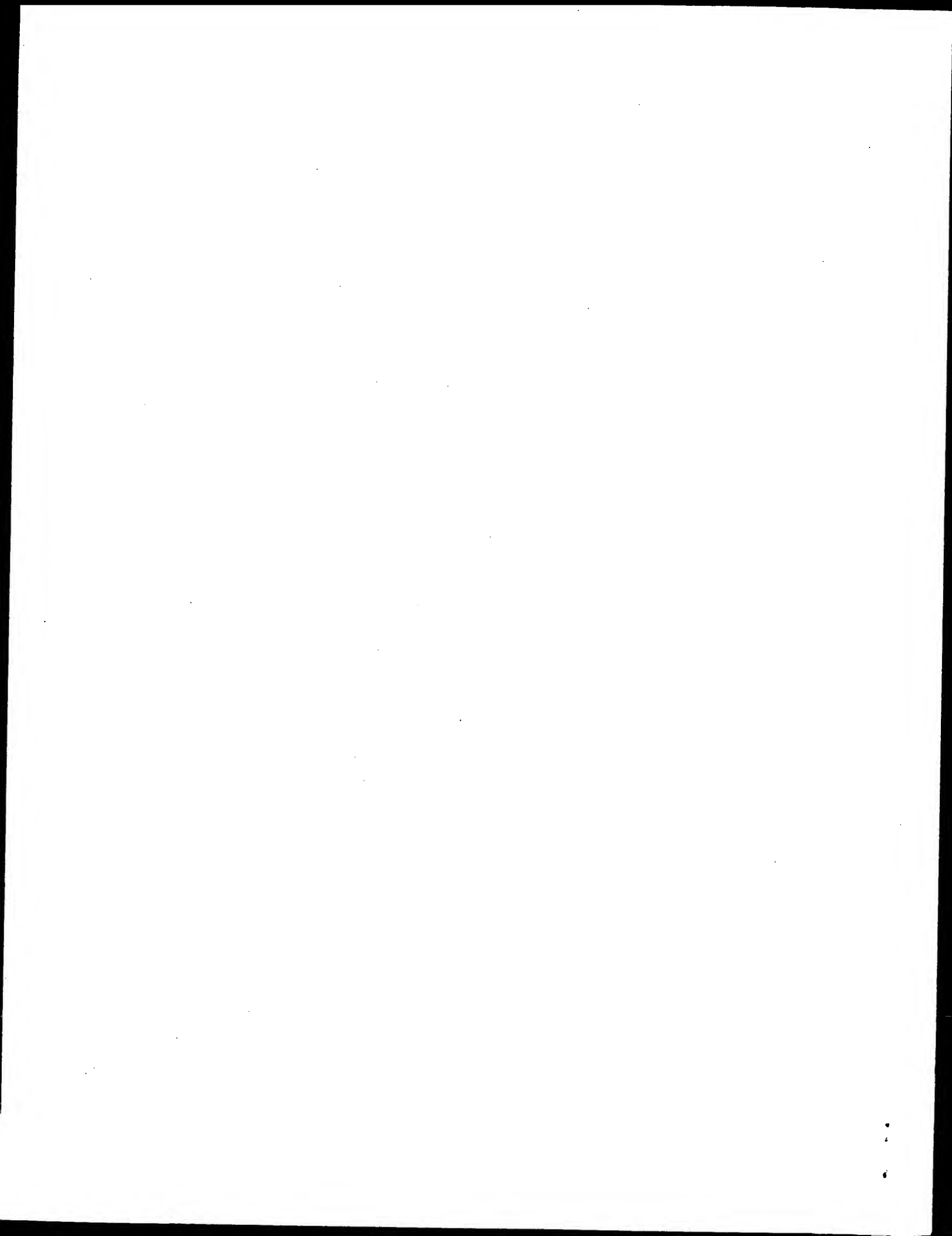
Matches 292; Conservative 69; Mismatches 127; Indels 4; Gaps 1;

```

Qy 16 RLLIRGRIVNDQSFYADYVHVEDGLIKOIGENLIYVG---GIHTIDAHGLMLVPGGVD 71
Db 6 RLLIRGRVYVNDSEVADYVHVEDGVVRLGHDLPPGGAPAGRLVLDRAKGLVLEGGID 65
Qy 72 VHTRLQMPVLGWTTPADDFCOGTAAALAGGTTMLDHFVDTGVSLLAAAYEQWRERADSA 131
Db 66 THTHMQFPFPGSRSIDDFHGTAAALSGGTTMIIDFAIPQKGGSLIEAFETWRSWADPKV 125
Qy 132 CCQYSLHVDITRWHESTKELEALVKEKGVNSFLVPMAYKDRCCQSDSQMYEYFISIRDL 191
Db 126 CCQYSLHVAVTWMSDOYKEEMKILVQDKGVNSFMFMAYKDYVVTDLLEYEAFSRCKEY 185
Qy 192 GALAQVHAENGDIIVEEBOKRLLELGTGPEGHVLSHPHEEVEAEAVYRAVTIAQANCPY 251
Db 186 GATAQVHAENGDLIAEGAKKMLALGITGPEGHELCRPEAEAEATLRAITIASAVNCPY 245
Qy 252 VTKVNSKAADAIAOAKRRGVVVEGEPITASLGTGDSHYWSKNWAKAAAFVTPPPVNDP 311
Db 246 IVHNSKSAKVIAADARRDGKVYGEPTAAASLGTGTHYWNKEWHHAHHVMGPPLRDP 305
Qy 312 TTADHLICLLSSGDLQVTSAGHCTTTAOKAVGKDNFALIPGEGTNGIEERMSVWKEKVA 371
Db 306 STPDFLMLNLANDLTTGTDCNTFCQKALGKDDFTKIPNGVNGVEDRMSVWEKGVH 365
Qy 372 SGKMDENEFVATSTNAAKIFNYPKRGVAVGSDADLVINPKATKIISAKTHNLNVEY 431
Db 366 SGKMDENEFVATSTNAAKIFNLYPRGRIAVGSDADIVWDPKGRITISAKTHHQAVNF 425
Qy 432 NIFEGVCEGAPAVVISOGRALEDGKMFVTPGAGRFVPRKTFPDFVYKRIKARNRLAEI 491
Db 426 NIFEGMVCHGVPLVTISRKVVYEAQVSVTAGDGKFIIPKPFAYEYIYKRIKQDRCTP 485
Qy 492 HGVPRGLYDGPV 503
Db 486 TPVERAPYKEV 497

```

Search completed: July 30, 2001, 11:42:21
Job time: 140 sec



.Tue Jul 31 13:08:20 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:37:26 ; Search time 16.35 Seconds
(without alignments)
2664.945 Million cell updates/sec

Title: us-09-367-496-8
Perfect score: 2985
Sequence: 1 MSFQKKIPRITSDRLIR.....RTAQKIMPPGGRSNTISLS 572

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2362	79.1	572	2 S58889	collapsin response
2	2343	78.5	572	2 S49985	dihydropyrimidinas
3	2339	78.4	572	2 JC5317	dihydropyrimidinas
4	2187	73.3	570	2 S55525	ulip protein - mou
5	2186	73.2	570	2 JC5318	dihydropyrimidinas
6	2151	72.1	572	2 JC5316	dihydropyrimidinas
7	1554	52.1	519	2 JC5315	dihydropyrimidinas
8	1548	51.9	519	2 S70581	hypothetical prote
9	1245	41.7	520	2 T20007	hypothetical prote
10	1122	37.6	544	2 T23968	dihydropyrimidinas
11	1074.5	36.0	479	2 H83590	unc-33 protein - C
12	837	28.0	471	1 JC2310	hypothetical prote
13	781.5	26.2	854	2 S33558	hypothetical prote
14	692.5	23.2	467	2 T28685	hypothetical prote
15	672	22.5	465	2 A65071	probable allantoin
16	672	22.5	465	2 F85942	allantoinase BH230
17	325.5	10.9	448	2 E75429	probable allantoin
18	317	10.6	438	2 E83938	probable dihydroor
19	311.5	10.4	453	1 G64782	allantoinase homol
20	296.5	9.9	430	2 B70959	probable dihydroor
21	295.5	9.9	446	1 C70016	dihydroorotase - M
22	295.5	9.9	454	2 A69017	dihydroorotase [im
23	259.5	8.7	422	2 D84403	dihydroorotase (EC
24	242	8.1	430	2 T45151	dihydroorotase pyr
25	241.5	8.1	422	2 C70370	hypothetical prote
26	233.5	7.8	428	2 B83967	pyrimidine synthet
27	231.5	7.8	2198	2 T20371	hypothetical prote
28	229.5	7.7	2242	2 A57541	hypothetical prote
29	229	7.7	296	2 H85550	hypothetical prote

ALIGNMENTS

RESULT 1

S58889

collapsin response mediator protein, 62K - chicken

C:Species: Gallus gallus (chicken)

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000

C:Accession: S58889

R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.

A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein

A:Reference number: S58889; MUID:95364923

A:Accession: S58889

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-572 <GOS>

A:Cross-references: EMBL:U17277; NID:9882146; PIDN:AAA93200.1; PID:9882147

C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 79.1%; Score 2362; DB 2; Length 572;

Best Loca: Similarity 76.4%; Pred. No. 8e-166;

Matches 436; Conservative 68; Mismatches 67; Indels 0; Gaps 0;

QY	1	MSFQKKIPRITSDRLIRGRIVNDQSFADVHVEDGLIKOIGENLIVPGGIHTDA	60
DB	1	MSYQKKIPRITSDRLIRKGGKIVNDQSFADVHVEDGLIKOIGENLIVPGVKVTKIA	60
QY	61	HSLMVLPGGVVHTRLQMPVLGTPADDFCQGTKAALAGGTTMILDRVPPDTGVSLAAY	120
DB	61	HSRMVLPGGIVHTRFQMPVQGTSAADDFCQGTKAALAGGTTMILDRVPPDTGVSLAAY	120
QY	121	EWRERADSAACDYSILHVDITRWHSIKLEALYKGVNSFLVFMAYKRCQSDSQ	180
DB	121	DQREWADSKSCDYSILHVDITRWHSIKLEALYKGVNSFLVFMAYKRCQSDSQ	180
QY	181	MYEFTSIIRDLGALAQVHAENGDIIVEEQKRLLELGTGPEGHVLSHPEVEAEVYRA	240
DB	181	IYEVLSVIRDIGTAQVHAENGDIIVEEQKRLLELGTGPEGHVLSHPEVEAEVYRA	240
QY	241	TIQAQNCPLYVTYKMSKGAADAIAQARRGVVYVGEPIITASLGDSHVSNNAKAAA	300
DB	241	TIQAQNCPLYVTYKMSKGAADAIAQARRGVVYVGEPIITASLGDSHVSNNAKAAA	300
QY	301	FVTSPPVNPDPPTADHITLCLLSSGDLQVTSAGCTFTTAQKAVGKDNFALIPCTNGIEE	360
DB	301	FVTSPPVNPDPPTADHITLCLLSSGDLQVTSAGCTFTTAQKAVGKDNFALIPCTNGIEE	360
QY	361	FMSVWEKCVASGKMDNEFVAVTSTNAKIFNFPYKRGVAVGSDADLVINPDKATKII	420
DB	361	FMSVWEKCVASGKMDNEFVAVTSTNAKIFNFPYKRGVAVGSDADLVINPDKATKII	420
QY	421	SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK	480
DB	421	SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPDFVYK	480

dihydroorotase (EC
dihydroorotase (EC
dihydroorotase (EC
allantoinase (EC 3
dihydroorotase (py
probable dihydroor
aspartate carbamoy
pyrimidine synthet
pyrimidine synthet
noncatalytic dihyd
dihydroorotase - D
urease (EC 3.5.1.5
dihydroorotase (EC

30 221.5 7.4 423 2 A64496
31 220 7.4 427 2 I40167
32 220 7.4 441 2 E86758
33 213 7.1 449 2 D82736
34 212 7.1 460 1 S48489
35 206.5 6.9 404 2 C75027
36 205 6.9 417 2 C71212
37 204.5 6.9 424 2 B56144
38 204 6.8 427 2 T48955
39 204 6.8 1481 1 OZDOP3
40 203 6.8 2225 1 A23443
41 200 6.7 445 2 B82954
42 199.5 6.7 423 2 G83595
43 197.5 6.6 416 2 E75437
44 196 6.6 569 2 C36950
45 195.5 6.5 428 1 DEBSO

Db 421 SAKTHNISLEYNIFEGMECRGSPVVISQKIVLELDGNLHVTEGSGRYIPKRPFFDFVYK 480
 QY 481 RIKARNRLAEIHGVPRLGYDGPVHEVWVPKPGSGAPARASCPGKISVPPVNLHOSGFS 540
 Db 481 RIKARSRLAELRGVPLGLYDGPVCEVSVTPKVTTPASSAKTSPAKQAPPPVNLHOSGFS 540
 QY 541 LSGSQADDDHARTAKOIMAPPGGRSNTISL 571
 Db 541 LSGAQIDDDNIPRTTQRIVAPPGGRANITSL 571

RESULT 2
 S49985
 dihydropyrimidinase-related protein 2 [similarity] - rat
 N:Alternate names: collapsin response mediator protein 2; TOAD-64
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 13-Jan-1995 #sequence_revision 12-May-1995 #text_change 02-Sep-2000
 C:Accession: A59280; S49985
 R:Minturn, J.E.; Fryer, H.J.; Geschwind, D.H.; Hockfield, S.
 J. Neurosci. 15, 6757-6766, 1995
 A:Title: TOAD-64, a gene expressed early in neuronal differentiation in the rat, is related to the human dihydropyrimidinase-related protein 2 (DRP-2)
 A:Reference number: A59280; MUID:96033765
 A:Accession: A59280
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-572 <MIN>
 A:Cross-references: GB:246882; NID:G599965; PIDN:CAA86981.1; PID:G599966
 A:Experimental source: strain Sprague-Dawley; clone TOAD64 (turned on after division, 64 hours after birth)
 A:Note: submitted to the EMBL Data Library, November 1994
 A:Note: in Genbank entry R10A064, release 113.0, the source is designated as Rattus rattus
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 78.5%; Score 2343; DB 2; Length 572;
 Best Local Similarity 75.8%; Pred. No. 2e-164;
 Matches 433; Conservative 70; Mismatches 68; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSRLIRGGIRVNDQSFYADVHVDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQGNIPRITSRLIRGGIRVNDQSFYADVHVDGLIKOIGENLIVPGGIHTIDA 60
 QY 61 HGLWLPGGVDVHTRQLMPVIGMTADDPCOGTKAALAGTTMIIDHVPDPTGVSLLAAY 120
 Db 61 HSRWIPGGIDVHTRQMPDQGMTSADDFQGTQKALAGTTMIIDHVPDPTGVSLLAAY 120
 QY 121 EOWREADSAACDYSYLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 DQREWADSKCCDYSYLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEIFSIIIRDLGALAQAQVHAENGDIIVEEOKRLELIGITGPEGHVLSHPPEEAEAVYRAV 240
 Db 181 IYEVLSVIRDLGALAQAQVHAENGDIIVEEOKRLELIGITGPEGHVLSHPPEEAEAVYRAV 240
 QY 241 TIAQANCPPLYTKVMSKGAADAIAQAKRGVNVVGEPIITASLTGDSHYHKNWAKAAA 300
 Db 241 TIANQTNCPPLYTKVMSKGAADAIAQAKRGVNVVGEPIITASLTGDSHYHKNWAKAAA 300
 QY 301 FVTSPPVNDPTDHLTCLSSGLDQVTSAGHCTFTTAAKAVGKDNFALLPEGTNGTEE 360
 Db 301 FVTSPPVNDPTDHLTCLSSGLDQVTSAGHCTFTTAAKAVGKDNFALLPEGTNGTEE 360
 QY 361 RMSWWEKCVASGKNDENEFVAVTSTNAKIFNYPKRGVAVGSDADLVINPKNATKII 420
 Db 361 RMSVWVAVGKNDENEFVAVTSTNAKIFNYPKRGVAVGSDADLVINPKNATKII 420
 QY 421 SAKTHNLNVEYNIFEGVBCRGAPAVISQGRVALEDGKMFVTPGAGRFVPRKTPDPVYK 480
 Db 421 SAKTHNSALEYNIFEGVBCRGAPAVISQGRVALEDGKMFVTPGAGRFVPRKTPDPVYK 480
 QY 481 RIKARNRLAEIHGVPRLGYDGPVHEVWVPKPGSGAPARASCPGKISVPPVNLHOSGFS 540
 Db 481 RIKARSRLAELRGVPLGLYDGPVCEVSVTPKVTTPASSAKTSPAKQAPPPVNLHOSGFS 540

QY 541 LSGSQADDDHARTAKOIMAPPGGRSNTISL 571
 Db 541 LSGAQIDDDNIPRTTQRIVAPPGGRANITSL 571

RESULT 3
 JC5317
 dihydropyrimidinase-related protein 2 - human
 N:Alternate names: collapsin response mediator protein 2
 C:Species: Homo sapiens (man)
 C>Date: 01-May-1997 #sequence_revision 23-Aug-1997 #text_change 02-Sep-2000
 C:Accession: JC5317; S58891; JC7150
 R:Hamejima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5317
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <BAM>
 A:Cross-references: DDBJ:D78013; NID:G1330239; PIDN:BAAL1191.1; PID:G1330240
 A:Experimental source: fetal brain
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58891
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <GOS>
 A:Cross-references: EMBL:U17279; NID:G1244399; PIDN:AAA93202.1; PID:G1244400
 R:Kitamura, K.; Takayama, M.; Hamejima, N.; Nakanishi, M.; Sasaki, M.; Endo, Y.; Take
 DNA Res. 6, 291-297, 1999
 A:Title: Characterization of the human dihydropyrimidinase-related protein 2 (DRP-2)
 A:Reference number: JC7150; MUID:20039612
 A:Accession: JC7150
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-572 <KIT>
 A:Cross-references: DDBJ:AB020764; PIDN:BA88422.1
 C:Comment: This protein is the homologue of chicken 62K collapsin response mediator p
 C:Genetics:
 A:Gene: GDB:DPYSL2; DHPRP2; DRP-2; CRMP2
 A:Cross-references: GDB:5496401
 A:Map position: 8p22-8p21
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 78.4%; Score 2339; DB 2; Length 572;
 Best Local Similarity 75.5%; Pred. No. 4e-164;
 Matches 431; Conservative 72; Mismatches 68; Indels 0; Gaps 0;

QY 1 MSFGKKSIPRITSRLIRGGIRVNDQSFYADVHVDGLIKOIGENLIVPGGIHTIDA 60
 Db 1 MSYQGNIPRITSRLIRGGIRVNDQSFYADVHVDGLIKOIGENLIVPGGIHTIDA 60
 QY 61 HGLWLPGGVDVHTRQLMPVIGMTADDPCOGTKAALAGTTMIIDHVPDPTGVSLLAAY 120
 Db 61 HSRWIPGGIDVHTRQMPDQGMTSADDFQGTQKALAGTTMIIDHVPDPTGVSLLAAY 120
 QY 121 EOWREADSAACDYSYLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 DQREWADSKCCDYSYLVHVDITRWHSIKEELEALVKRGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEIFSIIIRDLGALAQAQVHAENGDIIVEEOKRLELIGITGPEGHVLSHPPEEAEAVYRAV 240
 Db 181 IYEVLSVIRDLGALAQAQVHAENGDIIVEEOKRLELIGITGPEGHVLSHPPEEAEAVYRAV 240
 QY 241 TIAQANCPPLYTKVMSKGAADAIAQAKRGVNVVGEPIITASLTGDSHYHKNWAKAAA 300
 Db 241 TIANQTNCPPLYTKVMSKGAADAIAQAKRGVNVVGEPIITASLTGDSHYHKNWAKAAA 300

QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 QY 541 LSGSQADDDHARTAKIMAPPGRNSNITSLS 572
 Db 540 LSGSQADDDHARTAKIMAPPGRNSNITSLS 572

RESULT 5
 JC5318
 dihydroxyprimidinase related protein 3 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
 C:Accession: JC5318
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydroxyprimidinase and three related p
 A:Reference number: JC5315; MUID: 97128821
 A:Accession: JC5318
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-570 <HAM>
 A:Cross-references: DDBJ:D78014; NID:g1330241; PIDN:BAAL1192.1; PID:g1330242
 A:Experimental source: fetal brain
 A:Comment: This protein is the homologue of rat 64K turned on after division protein
 C:Genetics:
 A:Gene: GDB:IDPYSL3; DRP-3
 A:Cross-references: GDB:5885804
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 73.2%; Score 2186; DB 2; Length 570;
 Best Local Similarity 69.9%; Pred. No. 7e-153;
 Matches 400; Conservative 91; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSPOGKKSIPRITSDRLIRGGRIYVNDQSFYADYVHVEDGLIKQIGENLIVPGGIHTIDA 60
 Db 1 MSPOGKKSIPRITSDRLIRGGRIYVNDQSFYADYVHVEDGLIKQIGENLIVPGGIHTIDA 60
 QY 61 HGLMVLPGGVVDVHTRLOMPVLGTFPADDPCQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 HGLMVLPGGVVDVHTRLOMPVLGTFPADDPCQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 61 NGKMWIPGGIDVHTFQMPYKGTMTVDFFQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 NGKMWIPGGIDVHTFQMPYKGTMTVDFFQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEFTSIRDLGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 181 LYEFTCLGELGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 LYEFTCLGELGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 Db 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 QY 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 Db 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 QY 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 Db 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 QY 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 Db 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540

QY 301 FVTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 Db 301 FVTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 QY 541 LSGSQADDDHARTAKIMAPPGRNSNITSLS 572
 Db 540 LSGSQADDDHARTAKIMAPPGRNSNITSLS 572

RESULT 4
 S55525
 Ulip protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 02-Sep-2000
 C:Accession: S55525
 R:Byk, T.; Dobransky, T.; Cifuentes-Diaz, C.; Sobel, A.
 submitted to the EMBL Data Library, June 1995
 A:Description: Identification and molecular characterisation of Ulip, a putative mammal
 A:Reference number: S55525
 A:Accession: S55525
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-570 <BYK>
 A:Cross-references: EMBL:X87817; NID:g861058; PIDN:CAA61082.1; PID:g861059
 C:Genetics:
 A:Gene: Ulip
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 73.3%; Score 2187; DB 2; Length 570;
 Best Local Similarity 70.1%; Pred. No. 5.9e-153;
 Matches 401; Conservative 90; Mismatches 79; Indels 2; Gaps 2;

QY 1 MSPOGKKSIPRITSDRLIRGGRIYVNDQSFYADYVHVEDGLIKQIGENLIVPGGIHTIDA 60
 Db 1 MSPOGKKSIPRITSDRLIRGGRIYVNDQSFYADYVHVEDGLIKQIGENLIVPGGIHTIDA 60
 QY 61 HGLMVLPGGVVDVHTRLOMPVLGTFPADDPCQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 HGLMVLPGGVVDVHTRLOMPVLGTFPADDPCQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 61 NGKMWIPGGIDVHTFQMPYKGTMTVDFFQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 Db 61 NGKMWIPGGIDVHTFQMPYKGTMTVDFFQGTKAALAGGTTMILDHVPFDTGVSLLAAY 120
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 QY 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 EQRERADSAACDYSILHVDITRWHSIKEELEALYKGVNSFLVFMAYKDRCCSDSQ 180
 QY 181 MYEFTSIRDLGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 MYEFTSIRDLGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 181 LYEFTCLGELGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 LYEFTCLGELGALAQVHAENGDIIVEEQKRLLELITGPEGHVLSPHEVEAEAVYRAV 240
 QY 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 Db 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 QY 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 Db 241 TIAQANCPLVYKVMKSGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYWSKNWAKAAA 300
 QY 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 Db 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 QY 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 Db 301 FWTSPVNPDPDTADHLLTCLSSGDLQVTSACHTFTTAOKAVGKDNFALIPGFTNGIEE 360
 QY 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 Db 361 RMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKATKII 420
 QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 Db 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTPFPDFVYK 480
 QY 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540
 Db 481 RIKARNRLAEIHGVRGLYDGPVHEVMPAKPGCAPARASCPGKISVPPVRNLHQSGFS 540

Db 481 RIKARRKADHAYPRGMYDGPVDELTPKGTTPAGSARGSTRPN-PPVRNLHQSFGS 539
 Qy 541 LSGSQADHIAARRTAQKIMAPPGGRSNTLS 572
 Db 540 LSGTQVDEGV-RSASKRIVAPPGGRSNTLS 570

RESULT 6
 JC5316
 dihydropyrimidinase related protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 01-Dec-2000
 C:Accession: JC5316; S58890
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5316; MUID:97128821
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-572 <HAM>
 A:Cross-references: DBJ:D78012; NID:g1330237; PIDN:BAAL1190.1; PID:g1330238
 A:Experimental source: fetal brain
 R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
 Nature 376, 509-514, 1995
 A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein receptor
 A:Reference number: S58889; MUID:95364923
 A:Accession: S58890
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 64-503, 'H', 505-572 <GOS>
 A:Cross-references: EMBL:U17278; NID:g882148; PIDN:AAA93201.1; PID:g882149
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1994
 C:Comment: This protein is involved in nervous system development.
 C:Genetics:
 A:Gene: GDB:CRMP1; DRP-1; DPYSL1
 A:Cross-references: GDB:5585714
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 72.1%; Score 2151; DB 2; Length 572;
 Best Local Similarity 69.0%; Pred. No. 2.7e-150;
 Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;

Qy 1 MSFGKSPRIVSDRLIRGGRIVNDQSFYADVDEGLIKQIGENLIVPGGHTIDA 60
 Db 1 MSYGGKSPHITSDRLIKGGRIINDQSLYADVLEGLIKQIGENLIVPGGHTIEA 60
 Qy 61 HGLWLPVGGVYHTRLOMPVLGTPADFCQGTAAALAGGTTMILDHVFPDTGVSLLAAY 120
 Db 61 NGRMVPVGGIDVNTYLOKPSQGMTAADDFQGTAAALVGGTTMIDHVVPEPGSSLLTSF 120
 Qy 121 EWRERADSAACDYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQ 180
 Db 121 EKWHEADFKSCDYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQ 180
 Qy 181 MYEFTSLRDLGALQAQVHAENGDIIVEEQKRLLELGTITGPGHVLSPHEEAEAVYRAV 240
 Db 181 LYEATFLLGLGAVTLVHAENGDLIAQKRLLELGTITGPGHVLSPHEEAEAVYRAV 240
 Qy 241 TIAQANCPYVTKVMSKGAADATAQAKRGVVVFGPITASLTGDSHYWSKNWAKAAA 300
 Db 241 TIAGRINCPVITVTKVMSKGAADITALARKKPLVFGPITASLTGDSHYWSKNWAKAAA 300
 Qy 301 FVTPSPVNPDPPTADHLLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
 Db 301 FVTPSPVNPDPPTADHLLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 360
 Qy 361 RMSMWEKCVASGKNDENEFVAVTSTNAKIFNFKYPRKGVAVGSDADLVINWPKATKII 420
 Db 361 RMTVWMDKAVATGKMDENQFVAVTSTNAKIFNFKYPRKGVAVGSDADLVINWPKATKII 420

Qy 421 SAKTHNLNVEYIIFGVECRGAPAVVISOGRVALEDGKMFETPPGAGRFVPRKTFDFVVK 480
 Db 421 TAKSHKSAVEYIIFGMECHGSPVVISQGIKVFEDGNINKNKMGRIKPAFPEHLYQ 480
 Qy 481 RIKARNRLAEIHGVPRGLYDGVHVMVPAKPGSGAPARASCPCGKISVPPVRNLHQSFGS 540
 Db 481 RVKIRNKVFGQLQVSRGMVDPVVEVPATPKYATPAPSAKSSPSKHPPTIRNLHQSNGS 540

RESULT 7
 JC5315
 dihydropyrimidinase (EC 3.5.2.2) - human
 N:Alternate names: 5,6-dihydropyrimidine amidohydrolase; Hydantoinase
 C:Species: Homo sapiens (man)
 C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 18-Aug-2000
 C:Accession: JC5315
 R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
 Gene 180, 157-163, 1996
 A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins
 A:Reference number: JC5315; MUID:97128821
 A:Accession: JC5315
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-519 <HAM>
 A:Cross-references: DBJ:D78011; NID:g2339965; PIDN:BAAL1189.1; PID:g1330236
 A:Experimental source: liver
 C:Comment: This enzyme is Zn2+-metalloenzyme, and the second enzyme involved in uracil
 C:Genetics: dithyminase to N-carbamyl-beta-aminoisobutyrate. It catalyzes the hydrolysis of a vari
 A:Gene: GDB:DPYS; DHPase
 A:Cross-references: GDB:5885803
 C:Complex: homotetramer
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology
 C:Keywords: hydrolase

Query Match 52.1%; Score 1554; DB 2; Length 519;
 Best Local Similarity 59.3%; Pred. No. 1.9e-106;
 Matches 292; Conservative 69; Mismatches 127; Indels 4; Gaps 1;

Qy 16 RLLIRGGRIVNDQSFYADVDEGLIKQIGENLIVPG---GIHTDAHGLMVLPGVD 71
 Db 6 RLLIRGGRIVNDQSFYADVDEGLIKQIGENLIVPG---GIHTDAHGLMVLPGVD 65
 Qy 72 VHTRLQMPVLGTPADFCQGTAAALAGGTTMILDHVFPDTGVSLLAAYEOWRERADSA 131
 Db 66 THTHMQFPFGMSRSDIDFHQGTAAALSGGTTMIDFAIPKGGSLIEAFETWRSWADPKV 125
 Qy 132 CCDSYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQSYEIFSIIRDL 191
 Db 126 CCDSYSLHVDITRWHESTKEEALVKEGVNSFLVFMAYKDRQCSDSQSYEIFSIIRDL 185
 Qy 192 GALAQVHAENGDIIVEEQKRLLELGTITGPGHVLSPHEEAEAVYRAVYIAKQANCPY 251
 Db 186 GALAQVHAENGDLIAEGAKKMLALGITGPGHVLSPHEEAEAVYIAKQANCPY 245
 Qy 252 VTKVMSKGAADATAQAKRGVVVFGPITASLTGDSHYWSKNWAKAAAFVTPSPVNPDP 311
 Db 246 IVHVMKSAKVTADARRDGRKVVYGEPIAASLTGDTGYNNKHHAAHHVWGPPRLPDP 305
 Qy 312 TTADHLLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIEE 371
 Db 306 STPDFLMNLANDLLTCTDNCNTFCOKALGKDDFTKIPNGVNGVEDRMSVIEKGVH 365
 Qy 372 SKMDENEFVAVTSTNAKIFNFKYPRKGVAVGSDADLVINWPKATKII 431
 Db 366 SKMDENEFVAVTSTNAKIFNFKYPRKGVAVGSDADLVINWPKATKII 425
 Qy 432 NIFGVECRGAPAVVISOGRVALEDGKMFETPPGAGRFVPRKTFDFVVKIRKARNRLAEI 491

C;Accession: T20007

10

A:Residues: 1-544 <WIL>
A:Cross-references: EMBL:Z71266; PIDN:CAA95842.1; GSPDB:GN00019; CESP:R06C7.3
A:Experimental source: clone R06C7
C:Genetics:
A:Gene: CESP:R06C7.3
A:Map position: 1
A:Introns: 103/3; 135/2; 195/3; 255/3; 325/3; 418/2; 440/2; 522/3
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 37.6%; Score 1122; DB 2; Length 544;
Best Local Similarity 46.7%; Pred. No. 1.2e-74;
Matches 218; Conservative 79; Mismatches 136; Indels 34; Gaps 4;

QY 14 SDRLLRGRIVNDQSFYADV-----HVEDGLIKOI 45
DB 81 SPLPLVKNQTVNEDGKFRADVLTVDVSLQISCNFNFLNWTQYTKLHRH---REV 137
QY 46 GENLIVPGGIHTIDAHGLMVLPGVDVHTRLQMPVLGMPADDFCOGKKAALAGGTTMIL 105
DB 138 SPNITALPOTEVIDATDRIVIPGGIDPHTHMQMPYNGEVTKDDFLKGTAAVAGGTTMII 197
QY 106 DHVFPD--TGVSLAAYEQWRERADSACCDYSLHVDITRWHSIKEELEALVK--EKGVN 162
DB 198 DECCPDHNGESLIAGYNWRWSWADPKVCCDYGLSVATMRPETAQMAITSEFGVN 257
QY 163 SFLVMAYKDRQCSDSOMYEIFSIIRIDGALAQVHAENGDIIVEEQRKLELGITGPEG 222
DB 258 SFKFMAYENTLWVRDELFRQMCQALRALARVHCENGSVKEKEIDLLAKGVTPGEG 317
QY 223 HVLSPHEVEAVRYAVTIAKQANCLPYVTKVMSKGAADAIAQAKRGVVVFGPEPTAS 282
DB 318 HTQSPREEIEAETNRACVLAQAQANCPYIVHVKGAASAISHHRAQGGSTVFGPEPTAA 377
QY 283 LGTDSHYSKWNWAKAAAFVTPPNVDPPTADHLTCLLSSGDLQVTSAGHCTFTTAQA 342
DB 378 LALDGSYINEDWLHARYVMSPLSRDPTPELLMLLAAGELHGLTGDCNYDCRQS 437
QY 343 VGKONFALIPGTINGIERMSVWKECVASGKMDENEFVAVTSTNAKIFNFPYPRKGRVA 402
DB 438 LGKGNFTKIPNGINGVDRMSVWKEGVHSGIIDPRYVSTSTAAKIFNIYPRKGRIA 497
QY 403 VGSADLVINWPKATKIISAKTNLNVYINFEFVCGRGAAPVVISQ 449
DB 498 VGSADLVINWPKATKIISAKTNLNVYINFEFVCGRGAAPVVISQ 544

RESULT 11
H83590
dihydropyrimidinase PA0441 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83590
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:2043737
A:Accession: H83590
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <STO>
A:Cross-references: GB:AE004481; GB:AE004091; NID:99946293; PIDN:AA03830.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0441
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 36.0%; Score 1074.5; DB 2; Length 479;
Best Local Similarity 45.8%; Pred. No. 3.1e-71;
Matches 216; Conservative 81; Mismatches 166; Indels 9; Gaps 4;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGVDVHTRL 76
DB 3 LLIRGATVVTHEESYRADVLCAINGLETPSGCDVLDGGQYVLPGGIDPHTHM 62
QY 77 QMPVLGMPADDFCOGKKAALAGGTTMILDHVFPDPTGVSLAAYEQWRERADSACCDYS 136
DB 63 QLPFWGTVASEDFFGTAAGLAGGTTSIIDFIPNROSLLLEAFHTWRGWAQSA-ADYG 121
QY 137 LHVDITRWHSIKEELEALVKEGVNSFLVMAYKDRQCSDSOMYEIFSIIRIDGALAQ 196
DB 122 FHVATWNSDEVAREMGELVAHQHGWNSFKHFWAYKANATMAADDTLVASFERCLSELGAVPT 181
QY 197 VHAENGDIIVEEQRKLELGITGPEGHVLSHPEVEAEAVRYAVTIAKQANCLPYVTKVM 256
DB 182 VHAENGELVLFHQKLLAQGLTGPFAHPLSRPPOVEGEAASRAIRIAETLGTPLYLHIS 241
QY 257 SKGAADAIAQAKRGVVVFGPEPTASLGTDSHYSKWNWAKAAAFVTPPNVDPPTADH 316
DB 242 SREALDEITAVARAKGPQYVGEVLAGHLLDDSVYRHPDMATAAGYVMSPPFRP---VEH 297
QY 317 LTCL---LSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPGTINGIERMSVWKECVASG 373
DB 298 QEALWRGLQSGNLHTTATDHCFCACQKAMGRDDFSKIPNGTAGIEDRMALLDAGVNSG 357
QY 374 KMDENEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINWPKATKIISAKTNLNVYIN 433
DB 358 RLSMHFEVALTSTNTAKIFNFPKGAIRVAGADADLVMDPOGSRSTLSAATHHQRVDNI 417
QY 434 FEGVECRGAPVVISQGRVALEQKMFVTPGAGRFVPRKTFDFVYKRIKAR 485
DB 418 FEGTVRGIPSHSTISQKLLWAAGDLRALBPAGRYVERPAYDS-VYEVLRGR 468

RESULT 12
JC2310
dihydropyrimidinase (EC 3.5.2.2) - Bacillus stearothermophilus
N:Alternate names: hydantoinase
C:Species: Bacillus stearothermophilus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JC2310; PC2206
R:Mukohara, Y.; Ishikawa, T.; Watabe, K.; Nakamura, H.
A:Title: A thermostable hydantoinase of Bacillus stearothermophilus NS1122A: Cloning,
A:Reference number: JC2310; MUID:95036870
A:Contents: NS1122A
A:Accession: JC2310
A:Molecule type: DNA
A:Residues: 1-471 <MUK>
A:Cross-references: GB:S73773; NID:g688287; PIDN:AA060487.1; PID:g688288
A:Accession: PC2206
A:Molecule type: protein
A:Residues: 1-20 <MUK>
C:Comment: This enzyme is a metalloenzyme and the oligomeric structure is required for
C:Superfamily: allantoinase; Bacillus dihydroorotase homology
C:Keywords: hydrolase

Query Match 28.0%; Score 837; DB 1; Length 471;
Best Local Similarity 40.0%; Pred. No. 8.7e-54;
Matches 183; Conservative 79; Mismatches 192; Indels 4; Gaps 4;

QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGVDVHTRLQ 77
DB 4 LIRNGTIVTATDIYEADLLIQDKIAVGNRL-DESGAEVIDATGCVYFPGGIDPHTHLD 62
QY 78 MPVLGMPADDFCOGKKAALAGGTTMILDHVFPDPTGVSLAAYEQWRERADSACCDYS 137
DB 63 MPFGTVTKDDFESGTIAAAGGTTTIIDFCLTNKGEPLKKAETWNNKATGKAVIDYGF 122
QY 138 HVDITRWHSIKEELEALVKEGVNSFLVMAYKDRQCSDSOMYEIFSIIRIDGALAQ 197
DB 123 HLMISEITDDVLEELPKVTEEGITSEKFMAYKDVDFQADDGTLYRTLVAALGALVMV 182

Db 610 DVPLRG--APDRMIGALSTQPLVCTSGHRPVNSATRVAAKD-FAIAQKSGTGAERMA 666
 QY 364 MWKEKVASCKMDENEFVAVTSTNAAKIFENFYPRKGRVAVGSDADLVINWPKATKILIAK 423
 Db 667 VVHEHVRSGRIIDAMRFVAVTSTNAAKMFNPKGRIAGADADLVWDASGKRVLESS 726
 QY 424 THNUNVEYNIFEGVECKGAPAVVISQGRVALEDCKMEVTPGAGRFVPRKTFDFVYKRIK 483
 Db 727 RAOSSQENSMTDGLTVHSVVTATVIGKIAIYONGVEAREAPVAGFLRLSPNYPFLFSMVG 786
 QY 484 ARNKLAEIHGVPGLYDGPVHEVMVPAKPGSGAPAPARASCPCGKISVPPVRLNH-----535
 Db 787 QRDQFANVVER-----EASSQQ-----KPOQNGHHKSGDFD 821
 QY 536 -----OSGFSLSGSQADHIAHRTAQKIWAPPGBS 566
 Db 822 RNRTKVMESIDFGS-----AANRPRNPPGGRT 850
 RESULT 14
 T28685
 hypothetical protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T28685
 R:Parkhill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, January 1999
 A:Reference number: Z20512
 A:Accession: T28685
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-467 <PAR>
 A:Cross-references: EMBL:AL023496; PIDN:CAA18902.1
 C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match: 23.2%; Score 692.5; DB 2; Length 467;
 Best Local Similarity 33.8%; Pred. No. 3.7e-43;
 Matches 163; Conservative 84; Mismatches 192; Indels 43; Gaps 10;
 QY 14 SDRLLIRGRVNDQSFYADVHVEDGLIKOI-----GENLIVPGGIHTIDAH 61
 Db 2 SNTVIRGLVLTASDETHADVLIEDGRVAALAAATGPAEAFTAEV-----IDAS 53
 QY 62 GLMVLPGGVVDHTRLOMPVLGTPADDFCQGTKAALAGTTMTLDHVPDTPGVSLAAEY 121
 Db 54 GKYPVPGGVGDHTEMMPGGTYAATETGTTRAAANGTTTIVDFALQSVHSLREGLD 113
 QY 122 QWRERADSAACDYSILHVDITRWHSIKEELALVKGVNSFLVEMAYKDRCCSDSQM 181
 Db 114 AWHAKAEGNCAIDYGFHMIVSDVNQETLKEMDLLV-EEGVTFSKQFMAYPGVYSDQOI 172
 QY 182 YEIITSIRDLGALQAQVHAENGDIIVEEOKRLELGTGTPGEGHVLSPHEVEAEAVYAVT 241
 Db 173 LEAQRAAENGGLIMHAENGTAIDVLVEQALARGETDPRPHGEVVKALLAEATHRAIR 232
 QY 242 IAKQANCLYVTVKMSKGAADATAQAKRRGVVVFGE--PITASLCTDGGSHYKKNWA---296
 Db 233 LAQVAGAPLYVHVHVSATEVAELTRARDEGLPVFGTGPQYLFLLSTD-----NLAEPD 285
 QY 297 -KAAAFVTSPPVNDPTTADHLTCL---LSSGDLQVTSAGHCTFT--TAQAVGKDNFALI 351
 Db 286 FEGARYVCSTPLRP-----KEHOAALWRLTNDLVVSTDHCPFCFSGOKELGRGDSRI 341
 QY 352 PEGTNGIERMSVWEKCVASGKMDENEFVAVTSTNAAKIFENFYPRKGRVAVGSDADLV 411
 Db 342 PNGMPCVENRMDLL-HQAVVEGHIGRRRWRTEIACATPARMPGLTPKAGTAPACADIVV 400
 QY 412 WNPKATKIISAKTNLNVYNIFFGVECKGAPAVVISQGRVALEDCKMEVTPGAGRFVPR 471
 Db 401 YDPAEQVISAETHMNDYISAYEGRRTIGRVEIVLSRGEVPVVTREYTGKKGHAYTPR 460

QY 198 HAENGDIVEEOKRLELGLITGPEGRVLSHPEVEAEAVYAVTITAKOANCLYVTKVMS 257
 Db 183 HAENGVIDYLTAKALEDGHTDPIYHALTRPPELEGEATGRACQCLTELAGSLYVHVSC 242
 QY 258 KGAADATAQAKRRGVVVFGEPIITASIGTDSHWSKNWAKAAAFVTSPPVNDPTTADHL 317
 Db 243 QAQVEKTAEARNGLNVMGTCFQYLVLOSYLEKPNF-EGAKYVMSPPLR-EKWHQEVYL 300
 QY 318 TCLSSGLDQVTSAGHCTFT-TAOKAVGKDNFALIPGEGTNGIERMSVWEKCVASGRMD 376
 Db 301 WNALKNQQLTSGDQCSDFPKQKELGRGDTKIPNGGPILIEDRVSLIFSEGVYKGRIT 360
 QY 377 ENFVAVTSTNAAKIFENFYPRKGRVAVGSDADLVINWPKATKIISAKTNLNVYNIFFG 436
 Db 361 LNQFVDIVSTRIAKLFLPKKGTIAGADADLVIFDPTVERVISAETHMAMVDYNPFEF 420
 QY 437 VECRCAPAVVISQGRVALEDCKMEVTPGAGRFVPRKTF 474
 Db 421 MKVTGVPVSLRCGEFVYRDQFVKPGYGYVVRKRY 458
 RESULT 13
 S33558
 unc-33 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 07-Apr-1994 #sequence_revision 02-Jun-1994 #text_change 21-Jul-2000
 C:Accession: S33558; S33559; S24643; S24644
 R:Li, W.; Herman, R.K.; Shaw, J.E.
 Genetics 132, 675-689, 1992
 A:Title: Analysis of the Caenorhabditis elegans axonal guidance and outgrowth gene unc-33
 A:Reference number: S33558; MUID:93106371
 A:Accession: S33558
 A:Molecule type: DNA
 A:Residues: 1-854 <LIW1>
 A:Cross-references: EMBL:Z14148; NID:g6899; PIDN:CAA78520.1; PID:g6900
 A:Accession: S33559
 A:Molecule type: mRNA
 A:Residues: 1-854 <LIW2>
 A:Cross-references: EMBL:Z14146; NID:g6903; PID:g6904
 C:Genetics:
 A:Introns: 14/2; 68/1; 129/3; 151/3; 182/1; 255/3; 361/2; 660/1; 790/3
 A:Keywords: alternative initiators
 F:1-854/Product: unc-33 protein (long form) #status predicted <MARI>
 F:176-854/Product: unc-33 protein (intermediate form) #status predicted <MARI>
 F:332-854/Product: unc-33 protein (short form) #status predicted <MARI>
 Query Match: 26.2%; Score 781.5; DB 2; Length 854;
 Best Local Similarity 32.4%; Pred. No. 2.6e-49;
 Matches 187; Conservative 114; Mismatches 204; Indels 73; Gaps 11;
 QY 5 GKKSIPRITSDRLIRGGRIVNDQSFYADVHVEDGLIKOIGENLIVPGGIHTIDAHGLM 64
 Db 330 GEMSI-----LLVNAQIVNDDAIFAVIDIILEDGIIONAPNLEAPEGAEVLDAAKGL 382
 QY 65 VLPQGVVDHTRLOMPVLGTPADDFCQGTKAALAGTTMTLDHVPDTPGVSLAAEYQWR 124
 Db 383 ALPAGIDVYTVT-----DSSVDLSTGCKSAIAGTGTIVEVVRPRAESVSAVKRVK 437
 QY 125 ERAD-SAACDYSILHVDITRWHSIKEELALVKGVNSFLVEMAYKDRCCSDSQMVE 183
 Db 438 NOLEKSGISCHVALSVAITDFCE---QENSELYKNGINSFVL-----DGVSLTDOKLLE 489
 QY 184 IFSIIRDLGALQAQVHAENGDIIVEEOKRLELGTGTPGEGHVLSPHEVEAEAVYAVTIA 243
 Db 490 LFEHVKRLGALIRVVVPENKSVAMLEKMLKLGVTGPEFPQRPSPLEADRVSGVCVLG 549
 QY 244 KOANCLYVTVKMSKGAADATAQAKRRGVVVFGEPIITASIGTDSHWSKNWAKAAAFV 303
 Db 550 NIASCPISIVQVSSADSALAEKARASGALAHAEIASAAVTAADGALSQDLRFASAHLT 609
 QY 304 SPPVNDPTTADHLTCLSSGDLQVTSAGHCTFTTAQAVGKDNFALIPECTNGIERMS 363

Search completed: July 30, 2001, 11:41:24
Job time: 238 sec


```

68 GlyClyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
157 GCGGGTATTGATTCGACACACGATTTAGATATCCCTTTGGCGCACGGT 206
84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
207 GACAAAGGATGATTCGATCTGGACGATTCGGCGGCGATTTGGCGGAA 256
101 hrThrMetLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
257 CAACGACCATCATCGACTTTGTTTAAACGAATTAAGGGGACCATATAA 306
118 AlaAlaThrGluThrArgGluArgAlaAspSerAlaAlaCysCysAs 134
307 AAAGCGATGAACACTGGGCACACAAAGCGAAGGAAAGCGGTATTGA 356
134 pTyrSerLeuHisValAspLeuThrArgTrpHisGluSerLeuGlu 151
357 TTATGGCTTCCATTTAATCATTTAGCGAATTAACGATGACCTATTGA 406
151 LuLeuGluAlaLeuValLysGlyValAsnSerPheLeuValPhe 167
407 ACCTGCCAAAGTCAATCCGCAAGGAGGATAACATCCTTTAAAGTCTT 456
168 MetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
457 ATGGCGTATAAAGCGTATTCAGCGAGATGATGGAACGTTATACCGC 506
184 ePheSerLeuLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
507 GCTAGTGGCTGCCAAAGAACTTGGCGGCTGTGTCATGTTTCATCGGAA 556
201 snGlyAspLeuValGluGluGlnLysArgLeuLeuGluGlyLe 217
557 ATGGGATGCTGATTTACTTACGAAAGGAGCGCTTGGCGGAGGAAT 606
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
607 ACGAGCGCGATTTTACCATGCTTTAACGCGCCCTCCAGAGTAGAAGA 656
234 uAlaValTyrArgAlaValThrLeuAlaLysGlnAlaAsnCysProLeu 251
657 AGCAGCGCGCGCGCTGCTCAATTTGACAGAGCTTGGCGGTTTCAAC 706
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaLeuAlaGln 267
707 ACCTTGTTCACGTGACATGCGCAAGCGGTGGAAATAATTCACACGG 756
268 LysArgGlyValValPheGlyGluProIleThrAlaSerLeuGlu 284
757 CGCAATAAGGGTTGGATGTTGGGAGAAACGTGTCGCAATATCTTCT 806
284 yThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaP 301
807 TCTCGACCAATCG...TATTAGAAAGCCTGATTTGAAGCGCGAAT 853
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
854 ATGTTGGTCCCTCGCTCGT...GAAATAATGCAATCAAGATATTG 900
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHis 334
901 TGGATGCGCTGAAAGCGCGAGCTGCAAGCGCTTGGATCGGACCAATG 950
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
951 TTCATTGCTTAAAGGCGCAAGAAAGAACTTGGCAGAGAGATTTACTA 1000
350 euLeuProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
1001 AAATTCACACGGCGCGGATGTCGAGGATCGGTCGAGCATCTTTTC 1050
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAla 383

```

```

1051 AGTGAAGGGTTAAAAAGGAGGAATCACTAAATCAATTTGTCGATAT 1100
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGly 400
1101 TATGTCGACAGAATTCGAATTTGTCGGGTATTCCTCCGAGAAAGGAA 1150
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
1151 CGATCGCGGTAGTTTCAGACGACGACTTAGTCAATTTGACCCGGATATC 1200
417 ThrLysIleLeuSerAlaLysThrHisAsnLeuAsnValGluTyrAsn 433
1201 GAACGGGTGATTTGCGCGGAAACACCATATATGCGCGTCTGACTATA 1250
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerGln 450
1251 ATTTGAAGGAATGAAGTAACGGGTGAACCGGTATCGGTTCTGTGCAG 1300
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
1301 GCGAATTTGTTGTCGCTGATAAAACAATTTGTCGGAACACCGGTACGC 1350
467 ArgPheValProArgLysThrPhe 474
1351 CAATATTAAAGCTGCGTGTGTTT 1374

```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-602-656-2

```

seq_documentation_block:
; Sequence 2, Application
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Burtscher, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
; TITLE OF INVENTION: Production and Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; APPLICATION DATA:
; CURRENT APPLICATION NUMBER: US/08/602,656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-602-656-2

```


us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

alignment_scores:
  Quality: 826.00      Length: 458
  Ratio: 2.614        Gaps: 4
  Percent Similarity: 68.996      Percent Identity: 39.738

alignment_block:
US-09-367-496-8 x US-08-602-656-2 ..

Align seg 1/1 to: US-08-602-656-2 from: 1 to: 1383

18 LeuileArgGlyGlyValArgGlyValAlaAspAspGlnSerPheValAlaAs 34
  ::::::::::::::::::::|
10 ATAAATAAAATGGAACATGTTGTTACCCACACCGATACGTATGAAGCGGA 59
  ::::::::::::::::::::|
34 pValHisValGluAspGlyLeuLeuLysGlnIleGlyGluAsnLeuLeu 51
  ::::::::::::::::::::|
60 CTTGCTCATTAAGACGGAATAATGTCATGATAGGCCAACATTTA...G 106
  ::::::::::::::::::::|
51 alProGlyGlyLeuHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
  ::::::::::::::::::::|
107 AAGAAAAAGCGCTGAAGTGTATGATGCAAGGCTGTACGTATTTCCA 156
  ::::::::::::::::::::|
68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
  ::::::::::::::::::::|
157 GCGGTTATGATTCGACACGATTTAGATATGCGGTTTGGCGGCACCGT 206
  ::::::::::::::::::::|
84 rProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
  ::::::::::::::::::::|
207 GACAAAAGGATGATTTGCAATCTGGAACGATTCGCGCGCATTTGGCGGAA 256
  ::::::::::::::::::::|
101 hrThrMetLeuLeuAspHisValProAspThrGlyValSerLeuLeu 117
  ::::::::::::::::::::|
257 CAACGACCATCATCGACTTTGTTTAAACCAATAAAGGGGACCATTAATA 306
  ::::::::::::::::::::|
118 AlaAlaThrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAs 134
  ::::::::::::::::::::|
307 AAAGCGATGAACTTGGCAACAAAGGCAAGGAAAGCGGTATTGA 356
  ::::::::::::::::::::|
134 pTySerLeuHisValAspIleThrArgTrpHisGluSerIleLysGluG 151
  ::::::::::::::::::::|
357 TTATGCTTCCATTTATGATGATAGCGAAATACGGATGACGTATTAGAAG 406
  ::::::::::::::::::::|
151 luLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPhe 167
  ::::::::::::::::::::|
407 AGCTGCCAAAAGTCAATTCGCCAAGAGGGAATACATCTTTAAAGTGTT 456
  ::::::::::::::::::::|
168 MetAlaThrLysAspArgCysGlnCysSerAspSerGlnMetTyroGlu 184
  ::::::::::::::::::::|
457 ATGGCGTATAAAACGATTTTCAGCGAGATGATGGAACGTTATACCGCAC 506
  ::::::::::::::::::::|
184 ePheSerIleLeuArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
  ::::::::::::::::::::|
507 GCTAGTGGCTGCCAAAGAACTTGGCGGCTTGTGATGTTTCATGCGGAAA 556
  ::::::::::::::::::::|
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyLeu 217
  ::::::::::::::::::::|
557 ATGGGGATGATGATTTACTTACGAAAGAAAGCGCTTGGGAAGGGAAT 606
  ::::::::::::::::::::|
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
  ::::::::::::::::::::|
607 ACGGAGCGGATTTACCATGCTTTAAGCGGCTCCAGAAAGTAGAAGGAGA 656
  ::::::::::::::::::::|
234 uAlaValTrpArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 251
  ::::::::::::::::::::|
657 ACGACCGCGCGCGCTGTCAATTCAGACAGCTTCCCGGTTCAACAATTT 706
  ::::::::::::::::::::|
251 yValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
  ::::::::::::::::::::|
707 ACCTGCTTCCACGTGACATGCGCAAGCGGTGGAATAATTCACAACGC 756
  ::::::::::::::::::::|
268 LysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
  ::::::::::::::::::::|

```

```

757 CGCAATAAAGGGTTGGATGTGGGGAGAAACGCTGCCCAATATCTTGT 806
284 yThrAspGlySerHisTrpSerLysAsnTrpAlaLysAlaAlaAla 301
  ::::::::::::::::::::|
807 TCTCGACCAATCG...TATTTAGAAAAGCCTGATTTTGAAGCGCGCAAT 853
  ::::::::::::::::::::|
301 heValThrSerProProValAsnProAspProThrThrAlaAspHisLeu 317
  ::::::::::::::::::::|
854 ATGTTTGTGTCCTCCCTCGCT...GAAAAATGGCATCAAGAAGTATTG 900
  ::::::::::::::::::::|
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
  ::::::::::::::::::::|
901 TGGATGCGCTGAAAAACCGCCAGCTGCAACGCTGGATCGGACCAATG 950
  ::::::::::::::::::::|
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
  ::::::::::::::::::::|
951 TTCAATTTGACTTTAAAGCGCAAAAGAACTTGGCAGAGGAGATTTACTA 1000
  ::::::::::::::::::::|
350 euIleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
  ::::::::::::::::::::|
1001 AAATTCCAACCGCGCGGATGTCGAGGATCGGTCAGCATTTCTTTC 1050
  ::::::::::::::::::::|
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAla 383
  ::::::::::::::::::::|
1051 AGTGAAGGGTTAAAGAAAGAAAGTACAGTTAAATCAATTTGTCGATAT 1100
  ::::::::::::::::::::|
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTrpProArgLysGlyA 400
  ::::::::::::::::::::|
1101 TAGTCGACAAAGAAATGCGCAATTTGTCGGTTATTCCCGAGAAAAGAA 1150
  ::::::::::::::::::::|
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
  ::::::::::::::::::::|
1151 CGATCGCGGTAGTTCAGACGCGACTTAGTCATTTTACCCCGGATATC 1200
  ::::::::::::::::::::|
417 ThrLysIleIleSerAlaLysThrHisAsnLeuAsnValIleLysAsn 433
  ::::::::::::::::::::|
1201 GAAGCGGTGATTTCCGCGGAAACACACCATATGCGCGTCACTATAATG 1250
  ::::::::::::::::::::|
433 ePheGluGlyValGluCysArgGlyAlaProAlaValIleSerGluG 450
  ::::::::::::::::::::|
1251 ATTTGAAGGAATGAAGTAACCGGTGAACCGGTATCGGTCTCTGCGAG 1300
  ::::::::::::::::::::|
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
  ::::::::::::::::::::|
1301 CGCAATTTGTCGCGGTGATAACAATTTTCGGAACACCGGTACGGC 1350
  ::::::::::::::::::::|
467 ArgPheValProArgLysThrPhe 474
  ::::::::::::::::::::|
1351 CAATATTTAAAGGCTGGCTGTTT 1374
  ::::::::::::::::::::|

seq_name: /cgn2_6/ptodata/1/lna/58_COMB.seq:US-08-815-356-1

seq_documentation_block:
: Sequence 1, Application US/08815356
: Patent No. 5858759
: GENERAL INFORMATION:
: APPLICANT: Neal, Robert J
: APPLICANT: Griffin, Alison M
: APPLICANT: Scott, Miller O
: APPLICANT: Schatzman, Allan R
: APPLICANT: Gorham, Hazel C
: TITLE OF INVENTION: D-N-Carbamoyl-Amino Acid Amidohydrolase
: TITLE OF INVENTION: and Hydrantoinase
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation/Corporate
: ADDRESS: Intellectual Prop.
: STREET: P.O. Box 1539-UW2220
: CITY: King of Prussia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/815,356
; FILING DATE: 11-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,369
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervils, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P30433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5019
; TELEFAX: (610) 270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 891..1805
; US-08-815-356-1

```

alignment_scores:

```

Quality: 197.50 Length: 141
Ratio: 2.244 Gaps: 4
Percent Similarity: 62.411 Percent Identity: 39.716

```

alignment_block:

```

US-09-367-496-8 x US-08-815-356-1/rev ..

```

```

Align seg 1/1 to reverse of: US-08-815-356-1 from: 1 to: 1880

```

```

17 LeuLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAl 33
   ::::::::::::::::::::|
419 ATCATCATCAAGAACGGAACCATCGTAACCGCGGACGGGATTCCTCCGC 370
   |||||
33 aAspValHisValIleAspGlyLeuIleTyrGlnIleGly.GluAsnLeu 49
   |||||
369 CGATCTCGGAATCAAGATGCGAAGATCGCCAGATCGCGGGAACGTTCCG 320
   |||||
50 IleValProGlyGlyLeuHisThrIleAspAlaHisGlyLeuMetValLe 66
   |||||
319 GCCCGCGCGCGCGGA.....CAATCGACGCTCCGCGCGCTACGTTT 276
   |||||
66 uProGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyM 83
   |||||
275 TCGCGCGCATCGAGTTTCATACGATCTCGAGACGCTCGAGTTCAACA 226
   |||||
83 etThrProAlaAspPheCysGlnGlyThr..LysAlaAlaLeuAlaG 99
   |||||
225 CGCAGTCGCGCGCACATTCGCAACCGCGGCGTCTCGCGCGCC...TGTG 179
   |||||
99 lYgIlyThrThrMetIleLeuAspHisValPheProAspThrGlyValSer 115
   |||||
178 GCGGACGACGACCATCTGCTGATTTCTGCCAGCAGGACCGCGGCATAGC 129
   |||||
116 LeuLeuAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCy 132
   |||||
128 CTGAGGAGGCGGTGCCAAATGGAGCGCATGCCCGCGCAAGCTCGC 79
   |||||
132 scysAspTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleL 149
   |||||
78 GATCGACTACGGCTACCATATCATCTGCTCGATCCGACTGATAGCGCTGA 29
   |||||

```

```

149 ysGluGluLeuGluAlaLeu 155
   |||||
28 TCGAGAGCTAGAGGTACTG 9
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-732-242C-8
seq_documentation_block:
; Sequence 8, Application US/07732242C
; Patent No. 5298399
; GENERAL INFORMATION:
; APPLICANT: Uozumi, Takeshi; Masaki, Haruhiko;
; APPLICANT: Hidaka, Makoto; Nakamura, Akira;
; APPLICANT: Maeda, Michihisa; Yoneta, Yasuo
; TITLE OF INVENTION: Gene of Urease
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frishauf, Holtz, Goodman & Woodward, P.C.
; STREET: 600 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016-2088
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.72mb
; COMPUTER: IBM PC compatible (NEC PC-9801 RX)
; OPERATING SYSTEM: MS DOS
; SOFTWARE: ASCII Form
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07732,242C
; FILING DATE: 19910718
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JPN 2-210178
; FILING DATE: 10-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Herbert
; REGISTRATION NUMBER: 17081
; REFERENCE/DOCKET NUMBER: 910532/HG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)972-1400
; TELEFAX: (212)370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6131 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-07-732-242C-8

```

alignment_scores:

```

Quality: 196.00 Length: 616
Ratio: 0.737 Gaps: 26
Percent Similarity: 43.182 Percent Identity: 19.805

```

alignment_block:

```

US-09-367-496-8 x US-07-732-242C-8 ..

```

```

Align seg 1/1 to: US-07-732-242C-8 from: 1 to: 6131

```

```

1 MetSerPheGlnGlyLysLysSerIle.....Pr 10
   ::::::::::|
1106 GTAAAGTTTGGCGCGCAAGGTGATCCGAGATGGAATGGGCGCATCC 1155
   |||||
10 oArgIleThrSerAspArg.....LeuLeuIleArgGlyGlyArgI 24
   |||||
1156 TTTCGGCACAAGCGATGAAGCGTCGATCTCGTATTAAACAATGCGATTA 1205
   |||||
24 leValAsnAspAspGlnSerPheTyrAlaAspValHisValGluAspGly 40
   |||||
1206 TTGTTGATTACACAGGTATTATTAAAGCAGATATCGGCATAAAGATGGA 1255
   |||||

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

41  LeuTleLysGlnIleGlyGlu..... 47
1256 ATGATTCCTCCATAGAAAGCGGGAACCCCTTCTTAATGACGGGT 1305
48  AsnLeuIleValProGlyGlyLeuHisThrIleAspAlaHisGlyLeuM 64
1306 CGATATGGTGTGACGACGACAGCAAGATCATAGCCGAGAGGATGA 1355
64  etValLeuProGlyGlyValAspValHisThrArgLeuGlnMetProVal 80
1356 TTGTGACAGCGGAGGATAGATCATATTCATTTATTTGCCCT... 1402
81  LeuGlyMetThrProAlaAspPheCysGlnClyThrLysAlaAlaLe 97
1403 .....CAGCAATCGAACCCTCT 1422
97  uAlaGlyGly...ThrThrMetIleLeuAspHisValPheProAspThrG 113
1423 TGCATCGGTGTGACCATATGATTCGGGAGGAGAACAGGACCCGCTACAG 1472
113  lyValSerLeuLeuAlaAlaIleTyrGluGlnIleTyrArgLeuAlaAspSer 129
1473 GCACA.....AAT 1480
130  AlaAlaCysCysAspTyr-SerLeuHisValAspIleThrArgTyrHisG 146
1481 GCCACTACTGT.....ACACCGGGCCCTG 1506
146  uSerIleLysGluGluLeuAlaLeuValLysGlyLysValAsnS 163
1507 GAATATCCATCGTGTCTTCAAGCA.....GCCGAAG 1538
163  erPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAspSer 179
1539 AATTCGCGATAACTTGGGCTTTTAGGAAAGGAACTGTTCAGAT... 1585
180  GlnMetTyrGluIlePheSerIleLeuArgAspLeuGlyAlaLeuAlaG 196
1585 ..... 1585
196  nValHisAlaGluAsnGlyAspIleValGluGluGlnLysArgLeuL 213
1586 .....GAGGCTCCTTTAAGGACAAA 1607
213  euGluLeuGlyIleThrGlyProGluGlyHisValLeuSerHisProGlu 229
1608 TTGAACGGGCGGTGGGATTAAGCTTCAC.....GAA 1642
230  Glu.....ValGluAlaGluAlaValTyrArgAlaValThrIleAlaLy 244
1643 GATTGGGATCGACGGCGGCTGATTGATACATGTTTGAAGTGGCGGA 1692
244  sGlnAlaAsnCysProLeuTyrVal...ThrLysValMetSerLysGly. 259
1693 TCGATATGATGCAAGTATGCGATTATCATACAGACACTTAAATGAGGCG 1742
260  ....AlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValVal 274
1743 GATTGTGCGAGATACTTTGAAGCCATPAGACGTCGAGTATTCATACC 1792
275  PheGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTr 291
1793 TAT.....CATACAGAGGGCTGGCGGGGACAT..... 1822
291  pSerLysAsnTyrPalaLysAlaAlaPheValThrSerProProVala 308
1823 .GCTCGGATATATATAAAGCGCGGCTTCGCCAATATTTGCTCTCTT 1871
308  snProAspPro.....ThrThrAlaAspHisLeu 317
1872 CCACGAATCCAACTCGACCTTATATATCACTTTTGGAGAGCATTTA 1921
318  ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
1922 GATATGTTAATGGTTTCCACCACTAGACGCTAATATATCCAGAGGATAT 1971
334  sThrPheThrThrAlaGlnLysAlaValGlyLysAspAsn..... 347
1972 TGCTTTT.....GCCATTCACGCATACGAAAGAGAGACCATCGCGCGG 2015
348  ....PheAlaLeuIleProGluGlyThr 355
2016 AAGATGTTTATACATGATTAGGCGTTTTCAGCATGATTTCTGTGATTCA 2065
356  AsnGly.....IleGluGluArgMetSerMetValTyrPglulysCy 369
2066 CAGGCGATGGCGGAGTAGGAGAAGTATCTCTAGTGGCAACGCG 2115
369  s.....ValAlaSerGlyLysMetAspGluAsn..... 378
2116 TGACAAGATGAAAGCAAGAGAGGAGTTCACAAGAGACAACTGGTGG 2165
379  ....GluPheValAlaValThrSerThrAsnAla 388
2166 GAGACAACCTTCGTGTGAACGTTATATGCCAAATATACGATCAATCCG 2215
389  AlaLysIlePheAsnPheTyrProArgLysGlyArgValAlaValGlySe 405
2216 GCCATTCGTCATGCTATTCGGGATTTATCTGTTGAAGTGGGAA 2265
405  rAspAlaAspLeuValIleTyrAsnProLysAlaThrLysIleLeuSera 422
2266 ATTAGCTGATTTAGTGTGTGGATCTCT..... 2293
422  lalYThrHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGlu 438
2294 .....GCTTTTGTGTGAA 2311
439  CysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuG 455
2312 .....CCTGAACCTGGTCTTAAAGAGGAGGATGTTGCTTACAG 2349
455  uAsp..... 456
2350 CACTATGGAGATCCCAATGCCAGCATTCGCACACCGCGGTTTAT 2399
457  .GlyLysMetPheValThrProGlyAlaGlyArg..... 467
2400 ATCTCGCATGTTTGCAGCGAAGGAGATGCCAAATATCAACGCTCTATC 2449
468  ...PheValProArgLysThrPheProAspPheValTyrLysArgIleLy 483
2450 ACCTTTGTTCGAAAGCAGCGCTATGAAAGAGGATTCATCAACAGTTGGG 2499
483  sAlaArgAsnArgLeuAlaGluIleHisGlyVal..... 494
2500 TTTCGAAGAAAGGTGAACCAACGATCCATGGAATTCGAAATTCGACAAA 2549
495  ....ProArgGlyLeuTyrAspGlyPro 502
2550 AAGATTTAATTTGAACTAATAAACCACCAAAATTCAGTCGATCCTCAG 2599
503  ValHisGluValMetValProAlaLysProGlySerGlyAlaProAla 518
2600 ACATATGAAGTAAAGTAGAGCGTCAATTAGTAGACATGTGAACCGGCA 2647
seq_name: /cgn_2/ptodata/1/ina/5A_COMB.seq;us-08-967-513-1
seq_documentation_block:
; Sequence 1, Application US/08967513
; Patent No. 5783436
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: Of Use For Determination

```

48AsnLeuIleValProGlyGlyIleHisThr.... 57
957 ACCCGCATCCAGCCCAACGTCACCATCCCATCGCGCTCGCAGCGAA 1006
58 ...ileAspAlaHisGlyLeuMetValLeuProGlyGlyValAspValHI 73
1007 GTGATCCCGCCGAAGGAAATAATTGTACCGCGCGGGGATCATACCCA 1056
73 sThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspAspPheC 90
1057 TATTCACTGGATCTGTCGGCAG 1078
90 YSGlnGlyThrLysAlaAlaLeuAlaGlyGlyThrThrMetIleLeuAsp 106
1079 ..CAGCGGAAGAGCGCTGGTCTGGCGTGACCAACCATGTGTCGGCGC 1126
107 HisValPheProAspThrGlyValSerLeuLeuAlaAlaTyrGluGlnTr 123
1127 GGCACCGCGCCGCGCGCGGCACC 1150
123 pArgGluArgAlaAspSerAlaAlaCysAspTyrSerLeuHisVala 140
1151CATGCCACCACTGCACCCCGGC 1174
140 spIleThrArgTrpHisGluSerIleLysGluGluLeuGluAlaLeuVal 156
1175CCGTGGTATATCTCA 1189
157 LysGluLysGlyValAsnSerPheLeuValPheMetAlaTyrLysAspAr 173
1190CGCAT 1194
173 gCysGlnCysSerAspSerGlnMetTyrGluIlePheSerIleIleArgA 190
1195 GCTGCAGCGGCCGACAGCGTCGGGTCA 1223
190 spLeuGlyAlaLeuAlaGlnValHisAlaGluAsnGlyAspIleValGlu 206
1224 ATATCGGCTCTGGCAAGGAAACGTTTTCAGCGGATGCCCTCGCG 1273
207 GluGluGlnLysArgLeuLeuGluLeuGlyIleThrGlyProGluGlyHI 223
1274 GAGCAGGTGGCGGCAAGCGGTATTGGC 1300
223 svalLeuSerHisProGluGluValGluAlaGluAlaValTyrA 238
1301CTGAAGATCCAAGAGACTGGGCGCCACCCGCGCGCATCGACT 1346
238 rgAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyrVal 252
1347 GTCGTTAACGTCGCGGATGAATGGCATCCAGTCAGTCGCGCTGCACAGC 1396
253 ...ThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLys 268
1397 GACACCTGAATGAATCCGTTTTGTGGGAAGACACCTCGCGCCCATCGG 1446
268 sArgArgGlyValValPheGlyGluProIleThrAlaSerLeuGlyT 285
1447 CGGGCGCACCATCCACCTCCATACCGAAGGG 1481
285 hrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaPhe 301
1482 CGGGCGCGGCATGCGCGGACATCATCCGCTTCGCCCCAC 1525
302 ValThrSerProProValAsnProAspProThrThrAla 314
1526 CGGAACATTTGCGCTGCTCCACCAACCAACGCTGCCTACACCTCAA 1575
315AspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnV 328
1576 CACCATCATGAACATCTCGATATCTGATGGTCTGCCACCATCTGGACC 1625
328 alThrGlySerAlaHisCysThrPhe 336

```

1626 CGGACATCGCGGAGAGCTGGCCCTTTGCCGAGTCGCCATCGCCGGGAA 1675
      :::: ::::
337 ThrThrAlaGlnLysAlaValGlyLysAsp.....AsnPheAlaLeuIi 351
      ||| ||| :::: ||| ||| :::: |||
1676 ACCATCGCTGGGAAGAGCTGCTGCAGATCTCGCGCCCTTCTCGCTCAC 1725
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351 eProGluGlyThrAsnGly.....IleGluGluArgMetSerMetV 365
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1726 CTCCTCCGATTCGAGGCCATGGCGCGCTCGGGAGTGTATTCGCGCA 1775
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
365 aTriP6LysLysCys.....ValAlaSerGlyLysMetAspGlu 377
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1776 CTGGCAGGTGGCGCATGCGATGAAGTGCAGCGCGGAGCGCTGGCGAG 1825
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
378 .....AsnGluPheValAlaValTh 384
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1826 GAGACGGGGGATAACGACAACTTCGCGTGAAGCCCTACATCGCCCAATA 1875
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
384 rSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArgV 401
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1876 CACATCAACCGCGCGCTGACCCAGCGCATCGCACACAGTCGATCCA 1925
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
401 alAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThr 417
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1926 TTGAGTGGGTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1966
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
418 LysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePh 434
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1967 .....GCCTT 1971
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
434 eGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGlyA 451
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1972 CTTCCGGGTGAAA.....CCGGCCACCGTGATCAAGCGCGCA 2009
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
451 rgValAlaLeuGluAsp..... 456
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2010 TGATCGCCATCGCGCGCATGGCGGATCAATCAATCGCTCTATTCGACCCCG 2059
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPh 468
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2060 CAGCGCGTGCCTACCGCGCGATGTTGGCGCGGTGGCGAGCGCCGCCA 2109
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
468 eValProArgLysThrPhe.....ProAspPheValT 479
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2110 TCATCGCGCGCTCACCTTCCTGTCGAGCGCGCGCACCAATGGCGTTG 2159
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
479 yLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValPro 495
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2160 CCGAGCGGTGAACCTGGCGCGCGATGTCGCGGTGGTGAAGGTCGCGT 2209
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
496 ArgGlyLeuTyrAspGlyProValHisGluValMetValProAlaLysP 512
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2210 ACGGTGCAGAACCGCACATGTCGACACACAGTCGACGCGCTAACATCAC 2259
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
512 roGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSer 527
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2260 CGTCGAGCGCCGACCTATGAGGTGCGGTGGATGGCGGAACCTATCA 2306
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cgn2_6/ptodata/1/ins/5B_COMB.seq:US-08-687-645B-1
seq_documentation_block:
; Sequence 1, Application US/08687645B
; Patent No. 5846752
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: of Use For Determination
; TITLE OF INVENTION: of Urea
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway

```

176 ACCATCGTCGGGAAGACGTGCTGCAGCATCTCGGCCCTTCGCTCAC 1725

351 eProgluglyThrAsnGly.....IleGluluarGmetSerMetV 365
::: :::::::::::::::::::: ||| :::

1726 CTCCTCCGATTCCGAGGCATGGCCGCGTGGGGAAGATGATTCTCGCA 1775

365 alTrpGluLysCys.....ValAlaSerGlyLysMetAspGlu 377
:||||: :::::::::::::::::::: ||| :::
CTCGCAGGTGGC CATCGCATGAAGTG CAGCGCGAGCGCTGGCGGAG 1825

378AsnGluPheValAlaValTh 384

1826 GAGACCGGGGATAACGACAACCTCCCGTGAAGCGTACATGCCAAATA 1875

384 rSerThrAsnAlaAlaLysIlePheAsnPheTyProArgLysGlyArgV 401
::: ||| :::::::::::::::::::: ||| :::

1876 CACCATCAACCGCGCTGACCCACCGCATCGCACAGAAGTCGGATCCA 1925

401 alaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThr 417
::: |||||: :::::::::::::::::::: ||| :::
TTGAGGTGGTAAGCTGGCTGACTCGTGGTCTGGTCAACA..... 1966

418 LysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyAsnIlePh 434

1967GCCTT 1971

434 eGluGlyValGluCysArgGlyAlaProAlaValValIleSerGlnGlyA 451
CTTCGGCGGTGAAA.....CCGCCACCGTGATCAAAGCGGCA 2009

451 rgValAlaLeuGluAsp..... 456
::: ||||: :::::::::::::::::::: ||| :::

2010 TGATCGCCATCGCGCGGATGGCGGATATCAATGCTCTATTCCGACCCCG 2059

457GlyLysMet:PheValThrProGlyAlaGlyArgph 468
|||||: :::::::::::::::::::: ||| :::

1060 CAGCGCGTGCACTACCGCCCGATGTTTGGCGCGTGGCGAGCGCCGCCA 2109

468 eValProargLysThrPhe.....ProAspPheValT 479
||| |||||: :::::::::::::::::::: ||| :::

1110 TCATCGCGCGCTCACTTCCTTCGTCGCGAGCGCGGCAACGCAATGGCGTTG 2159

479 yrLysArgIleLysAlaArgAsnArgLeuAlaGluIleHisGlyValPro 495
:||||: :::::::::::::::::::: ||| :::
CGAGCGGCTGAACCTCGGCAGCGCATCGCGGTGGTGAAGAGCTCGCGT 2209

496 ArgGlyLeuTyAspGlyProValHisGluValMetValProAlaLysP 512

1210 ACGGTGCAAGAACCGCATGGTGCACAACAGCTGCAGCCTAACATCAC 2259

512 roGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSer 527
:||||: ||| :|||: :::::::::::::::::::: ||| :::
CGTCAGCGCCGACNCTTATGAGTGGGGTGGATCGAGCACTATCA 2306

seq_documentation_block:

GENERAL INFORMATION:

APPLICANT: Sauerba
APPLICANT: Ferrar

TITLE OF INVENTION:

:	TITLE OF INVENTION:
:	TITLE OF INVENTION:

CORRESPONDENCE ADDRESS

ADDRESSEE: Dunne
STREET: 1300 T s

10

Tue Jul 31 13:08:19 2001

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER: IBM PC compatible
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2619 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 31..36
OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
OTHER INFORMATION: sequence."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 756..759
OTHER INFORMATION: /standard_name= "Shine-Dalgarno"
OTHER INFORMATION: sequence."
US-08-467-822-19

alignment_scores:
Quality: 136.50 Length: 553
Ratio: 0.576 Gaps: 24
Percent Similarity: 42.857 Percent Identity: 19.711

alignment_block:
US-09-367-496-8 x US-08-467-822-19 ..

Align seg 1/1 to: US-08-467-822-19 from: 1 to: 2619

1 MetSerPheGlnGlyLysSerile..... 9
852 ATCAAAATTTGGGGGGTAAACATATCGTGATGGGATGAGTCAACAA 941
10ProArgIleThrSerAspArgLeuIleArgGlyGlyArgIleV 25
942 TAGCCCTAGCTCTTATGAATATAGATTGGTCTACTAACGCCCTCAT 991
25 alAspAspGlnSerPheTyrAlaAspValHisValcIuAspGlyLeu 41
992 TGGACTATACGGCATTTACAAAGCCGACATTGGGATTAAAGACGGCA 1041
42 IleLysGlnIleGlyClu..... 47
1042 ATTGCAGGCATTGGCAAGCGGCAATAGGACATGAGATGGCGGTAGA 1091

48AsnLeuIleValProGlyGlyIleHisThrIleAspAlaHisGlyL 63
1092 TAATAATCTTTGGTAGCTCTGTACAGAGGCTTTGGCAGCTGAGGGCT 1141
63 euMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMetPro 79
1142 TGATTGTAACCGCTGGTGGCATCGATACGATATTCACCTTATCTCTCC 1191
80 ValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaAl 96
1192CAACAAATCCCTACTGTC 1208
96 aLeuAlaGlyCly...ThrThrMetIleLeuAspHisValPheProAsp 112
1209 TTTTGGCAGCGGGTTACAAACCATGATGGAGGAGCACAGGACCTCGG 1258
112 hrGlyValSerLeuLeuAlaAlaTyrGluGlnTrpArgGluArgAlaAsp 128
1259 ATGGCAGCAATGCG..... 1272
129 SerAlaAlaCysCysAspTyrSerLeuHisValAspIleThrArgTrpHi 145
1273ACCACCATCATCTCCCGGACG 1292
145 sGluSerIleLysGluGluLeuAlaLeuValLysGlyLysGlyVala 162
1293 CGCTAATCTAAAGATGTTGGTGGCA...GCCGAGAAATACGCCATGA 1339
162 snSerPheLeuValPheMetAlaTyrLysAspArgCysGlnCysSerAsp 178
1340 AT.....CTAGGCTTTTGGCT...AAGGGGAATGTCTTACGACCC 1380
179 SerGlnMetTyrGluIle.....PheSerIleIleLeu 189
1381 TCTTTACGGATCAGATTGACAGCGGGCGATTGTTTAAATCCACGA 1430
189 gAspLeuGlyAlaLeu...AlaGlnValHisAlaGluAsnGlyAspIleV 205
1431 ACACCTGGGGAAGCACACCTGCAGCTATTAC..... 1461
205 alGluGluGlnLysArgLeuLeuGluGlyIleThrGlyProGlu 221
1461 1461
222 GlyHisValLeuSerHisProGluValGluAlaGluAlaValTyrAr 238
1462 ...CACTGCCTCAATGTCGCCGATGAATACGATGTCAAAGTGCTATCCA 1508
238 gAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyrValThrLysV 255
1509 CACCGATACCTT...AACGAGCGGGCTGT..... 1536
255 alMetSerLysGlyAlaAlaAlaIleAlaGlnAlaLysArgArgGly 271
1537GTAGAAGACACCTAGAGCGGATTCGCCGGCGCAC 1572
272 ValValValPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySe 288
1573 ATCCATACCTTC.....CACACTGAAGGGCTGGGGTGG 1607
288 rHisTyrTrpSerLysAsnTrpAlaLysAlaAlaPheValThrSerP 305
1608 ACAC.....GCTCCAGATGTTATCAAAATGGCAGGGGAATTTAACATTC 1651
305 roProValAsnProAspPro.....ThrThrAla 314
1652 TACCGCGCTCTACTAACCCGACCATTCCTTTACCAAAAAACACTGAAGCC 1701
315 AspHisLeuThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySe 331
1702 GAGCATGGGACATGTTAATGCTGTGCCACCACCTTGGATAAAAGTCAAA 1751


```

331 rAlaHisCysThrPheThrThra.....GlnLysAlaValG 344
1752 GGAAGATGTCAGTTTCGGGATTCGAGGATTCGCCCAAACTATCGCG 1801
344 lYlAspAsn.....PheAlaLeuIleProGluGly 354
1802 CTGAAGACCAACTCCATGATGATGGGATCTTTCTATCACCAGCTCCGAC 1851
355 ThrAsnGly.....IleGluGluArgMetSerMetValTIPsluLy 368
1852 TCTCAGGCTATGGGACGCGTAGCGGAGTGATCACACGCACTTGCAGAC 1901
368 sCysValAlaSer.....GlyLysMetAspGluAsnGlu.... 379
1902 ACAGACAAAAACAAAAGAGTTTGGCGCTTGAAGAGGAAAAAGCG 1951
380 .....PheValAlaValThrSerThrAsn 387
1952 ATAAGCACAACCTCCGATCAACGCTACATCTCTAATACACCATCAAC 2001
388 AlaAlaLysIlePheAsnPheTyrProArgLysGlyArgValAlaValG 404
2002 CCCGGATCGCGCATGGGATTTCTGACTATGTGGCTCTGTGGAAGTGG 2051
404 yserAspAlaAspLeuValIleTTPAsnProLysAlaThrLysIleIles 421
2052 CAATACCGCGACCTGCTGCTTGGAGTCG..... 2082
421 erAlaLysThrHisAsnLeuAsnValGluTyrAsnIlePheGluGlyVal 437
2083 .....GCTTCTTGGCAT 2097
438 GlucYargGlyAlaProAlaValIleSerGlnGlyArgValAlaLe 454
2098 AAG.....CCAAATATGATTATTAAAGGCGGATTTATTGCGCT 2135
454 uGluAsp..... 456
2136 CTCTCAAAATGGCGATGCCAATCGCTATTCACCACCCCTCAGCCGCT 2185
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPheValProArg 471
2186 ATTACCGTGAATGTTTGGACACCATGGGAAAAACAAATTCGACACCAAT 2235
472 LysThrPhe 474
2236 ATCATTTTC 2244

```

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq.US-08-432-697-19

```

seq_documentation_block:
; Sequence 19, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2619 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..36
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno
; OTHER INFORMATION: sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 756..759
; OTHER INFORMATION: /standard_name= "Shine-Dalgarno
; OTHER INFORMATION: sequence."
; US-08-432-697-19

```

alignment_scores:

```

Quality: 136.50      Length: 553
Ratio: 0.576         Gaps: 24
Percent Similarity: 42.857 Percent Identity: 19.711

alignment_block:
US-09-367-496-8 x US-08-432-697-19 ..

Align seq 1/1 to: US-08-432-697-19 from: 1 to: 2619

1 MetSerPheGlnGlyLysSerIle..... 9
892 ATCAAAATTTGGGCGGTAAACCTATCCGTATGGGATGAGTCAACCAA 941
10 ....ProArgIleThrSerArgLeuLeuIleArgGlyGlyArgIleV 25
942 TAGCCCTAGCTCTTATGAATTAGATTGGTGTCTACCTAACGCCCTCAT 991
25 alAsnAspAspGlnSerPheTyrAlaAspValHisValGluAspGlyLeu 41
992 TGGACTATACGGCATTTACAAAGCCGACATTTGGATTAAAGACGCGCA 1041
42 IleLysGlnIleGlyGlu..... 47
1042 ATTCAGCGCATTTGGCAGGCGAGCAATAAGGACATGCAAGATGCGCTA 1091
48 ....AsnLeuIleValProGlyGlyIleHisThrIleAspAlaHisGlyL 63
1092 TAATAATCTTTGCTAGTCTCTACAGAGGCTTTGGCAGCTGAGGGCT 1141
63 euMetValLeuProGlyGlyValAspValHisThrArgLeuGlnMetPro 79
1142 TGATTGTAACCGCTGGTGGCATCGATACCATTTTACTTTTCTCTCC 1191
80 ValLeuGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaAl 96
1192 .....CAACAAATCCCTACTGC 1208
96 aLeuAlaGlyGly...ThrThrMetIleLeuAspHisValPheProAspT 112

```

Tue Jul 31 13:08:19 2001

```
1209 TTTTCCAGCGGGTTTACAACTGATGTTGGAGGAGGCACAGGACCTGGCG 1258
112 hrGlyValSerLeuLeuAlaAlaTyTrpArgGluArgAlaAsp 128
1259 ATGGCAGGAATGG..... 1272
129 SerAlaAlaCysAspTyTrpSerLeuHisValAspIleThrArgTrpHi 145
1273 .....ACCACCATCCTCCCGGAGC 1292
145 sGluSerIleGlyGluLeuGluAlaLeuValGlyGlyVala 162
1293 CGCTAATCTAAAGTAGTGTGGTGCA...GCCGAGAAATACGCCATCA 1339
162 snSerPheLeuValPheMetAlaTyTrpIleAspArgCysGlnCysSerAsp 178
1340 AT.....CTAGGCTTTTGGCT...AAGGGAATGTCTTACGAACCC 1380
179 SerGlnMetTyTrpGluLeu.....PheSerIleIleAr 189
1381 TCTTTTACGCGATCAGATTGAACGAGCGCGGATGTTTAAATCCACGA 1430
189 gAspLeuGlyAlaLeu...AlaGlnValHisAlaGluAsnGlyAspIleV 205
1431 AGACTGGGAGGACACACCTGCGAGCTATTCAC..... 1461
205 aiGluGluGluGlnLysArgLeuGluLeuGlyIleThrGlyProGlu 221
1461 ..... 1461
222 GlyHisValLeuSerHisProGluGluValGluAlaGluAlaValTrAr 238
1462 ...CACTGCTCATGTGCCGATGAATACGATGTGCAAGTGGTATCCA 1508
238 gAlaValThrIleAlaLysGlnAlaAsnCysProLeuTyValThrLysV 255
1509 CACCGATACCTT...AAGGAGCGGGGTGT..... 1536
255 alMetSerLysGlyAlaAlaAspAlaIleAlaGlnAlaLysArgGly 271
1537 .....GTAGAGACACCTAGAGCGGATTCGGCGGCACCC 1572
272 ValValValPheGlyGluProIleThrAlaSerLeuGlyThrAspGlySe 288
1573 ATCATACCTTC.....CACACTGAAGGGGCTGGGGGTGG 1607
288 rHisTyTrpSerLysAsnTrpAlaLysAlaAlaPheValThrSerP 305
1608 ACAC.....GCTCCAGATCTTATCAAAATGGCAGGGGAATTTACATTC 1651
305 roProValAsnProAspPro.....ThrThrAla 314
1652 TACCGGCTCTACTAACCAGCATTCTTTCCACCAAAACACTGAAGCG 1701
315 AspHisLeuThrCysLeuSerSerGlyAspLeuGlnValThrGlySe 331
1702 GAGCAGATGGACATGTTAATGGTGTGCCACCCTTGGATAAAGATCA 1751
331 rAlaHisCysThrPheThrAla.....GlnLysAlaValG 344
1752 GGAAGATGTGAGTTGGCGATTCGAGGATTCGCCCAAACTATCGCG 1801
344 lLysAspAsn.....PheAlaLeuIleProGluGly 354
1802 CTGAAGACCAACTCCATGACATGGGGATCTTTTCTATCACCAGCTCCGAC 1851
355 ThrAsnGly.....IleGluGluArgMetSerMetValTrpGluLy 368
1852 TCTCAGGCTATGGACGCGTAGCGGTGATCATCACGCACTTGGCAGAC 1901
368 sCysValAlaSer.....GlyLysMetAspGluAsnGlu..... 379
```

```
1902 AGCAGACAAAAAAGAGTTTGGCGCTTGAAGAGGAAAGGCG 1951
380 .....PheValAlaValThrSerThrAsn 387
1952 ATAACGACAACTTCCGATCAAAACGCTACATCTCTAAATACACCATCAAC 2001
388 AlaAlaLysIlePheAsnPheTyTrpArgLysGlyArgValAlaValG1 404
2002 CCCGGGATCGCATGGGATTTCTGACTATGTGGGCTCTGTGGAAGTGGG 2051
404 ySerAspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleS 421
2052 CAATACGCGACCTCGTCTTGGAGTCG..... 2082
421 erAlaLysThrHisAsnLeuAsnValGluTyTrpAsnIlePheGluGlyVal 437
2083 .....GCTTTCTTTGGCAT 2097
438 GluCysArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLe 454
2098 AAG.....CCCAATATGATTATAAGGCGGATTTATTGCGCT 2135
454 uGluAsp..... 456
2136 CTCTCAATGGCGCATGCCAATGCGTCTATTCCACCCCTCAGCCCGTCT 2185
457 .....GlyLysMetPheValThrProGlyAlaGlyArgPheValProArg 471
2186 ATTACCGTGAAATGTTTGGACACCATGGGAAACAAATTCGACACCAAT 2235
472 LysThrPhe 474
2236 ATCACTTTC 2244
seq_name: /cyn2_6/ptodata/1/ina/5B_COMB.seq:US-08-920-095-1
seq_documentation_block:
; Sequence 1, Application US/08920095
; Patent No. 5837240
; GENERAL INFORMATION:
; APPLICANT: Cynthia K. Lee et al.
; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE
; NUMBER OF INVENTIONS: 3
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,095
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,041
; FILING DATE: 28-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
```

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-920-095-1

alignment_scores:
Quality: 136.50 Length: 632
Ratio: 0.493 Gaps: 29
Percent Similarity: 43.829 Percent Identity: 19.146
alignment_block:
US-09-367-496-8 x US-08-920-095-1 ..
Align seg 1/1 to: US-08-920-095-1 from: 1 to: 2735

```

17 LeuLeuIleArgGlyGlyArgIleValAlaAsnAspGlnSerPheTyrAla 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 TTAATTATCACTAACGCTTTAATCGTGGATTACACCGGTATTATATAAGC 1072
33 aspValHisValGluAspGlyLeuLeuLysGlnIleGlyGlu..... 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1073 GGATATTGTTAAGATTAAAGATGGCAAAATCGTCGCAATTGGTAAAGCGGTA 1122
48 .....AsnLeuIleValProGlyGly 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 ACAAGACATGCAAGATGCGTTAAACAACTTTAGCGTAGGTCCTGCT 1172
55 IleHisThrIleAspAlaHisGlyLeuMetValLeuProGlyGlyValas 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1173 ACTGAAGCCTTAGCGCGGTGAAGTTTGATCGTAACGGCTGGTGATTGA 1222
71 pValHisThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspA 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 CACACACATCCACTTCATTTCCACC..... 1247
88 spPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly...ThrThrMet 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1248 .....CAACAAATCCCTACAGCTTTTGCAACGCGTGAACACCATG 1289
104 IleLeuAspHisPheProAspThrGlyValSerLeuLeuAlaAlaTy 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1290 ATTGGTGTGGAACCGTCTCGTGTGAGGCACTAATGCG..... 1328
120 rGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAspTyrSerL 137
1328 ..... 1328
137 euHisValAspIleThrArgTrpHisGluSerIleLysGluLeuGlu 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1329 .....ACTACTATCACTCCAGGCGAGAAATTTAAATGGATGCTCAGA 1373
154 AlaLeuValLysGluLysGlyValAsnSerPheLeuValPheMetAlaTy 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1374 GCG...GCTCAACAATATCTATGANT..... 1397
170 rLysAspArgCysGlnCysSerAspSerGlnMetTyrGluIlePheSerI 187
1397 ..... 1397
187 leIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnGlyasp 203
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 .....TTAGTTTCTTGGCTAAAGGTAAAGCTTCAACGATCGC 1436
204 IleValGluGluGlnLysArgLeuLeuGluLeuIleThrGly... 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1437 ACCTTAGCGGATCAA.....ATTGAAGCGGTGGCGATTGGCTT 1474
220 .....ProGlu.....GlyHisValL 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1475 TAAATTCACGAAGACTGGGCGACCACTCTTCGCAATCAATCATCGCT 1524
225 euSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThr 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1525 TAGATGTTGGGACAAATAGATGTGCAAGTCGCTATCCACACACACT 1574
242 IleAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerLys 258
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1575 TTG...AATGAAGCGGTGT..... 1592
258 sGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValp 275
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1593 .....GTAGAAGACTATGCGTGTATTGCTGGAGCGCACTATGCACACTT 1638
275 heGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrp 291
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1639 TC.....CACACTGAAGCGGTGGCGCGGACAC..... 1667
292 SerLysAsnTrpAlaLysAlaAlaPheValThrSerProProValas 308
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1668 GCTCTGATATTATAAGTAGCGGTGAACACAAACATTTCTCCGCTTC 1717
308 nProAspPro.....ThrThrAlaAspHisLeuT 318
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1718 CACTAACCCCCACCCTTCACCGCTGAATACAGAAGCAGACGACATGG 1767
318 hrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCys 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 ACATGCTTATGTTGTCGCCACCTTGGATAAAGCATTAAGAAGATGTT 1817
335 ThrPheThrThrAla.....GlnLysAlaValGlyLysAspAs 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1818 CAGTTGCTGATTCAAGGATCCGCCCTCAACCATTTGGCGGTGAAGACAC 1867
347 n.....PheAlaLeuIleProGluGlyThrAsnGly. 357
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1868 TTTGCATGACATGGGATTTTCTCAATCACCAGTTCTGACTCTCAAGCA 1917
358 .....IleGluGluArgMetSerMetValTrpGluLysCysValAla 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1918 TGGCGCGTGTGGGTGAAGTATACACTAGAACTTGGCAACACAGCTGACAA 1967
372 Ser.....GlyLysMetAspGluAsnGlu..... 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1968 AACAGAAGAAATTTGGCGCTTGAAGAAGAAAGGCGGTAACGACAA 2017
380 .....PheValAlaValThrSerThrAsnAlaAlaLysI 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2018 CTTAGGATCAACCGCTACTGTCTAAATACACCATTAAACCGCGATCG 2067
391 lePheAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAla 407
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2068 CTCATGGGATTAGCGAGTATGAGTTTCAGTAGAAGTGGGCAAGTGGCT 2117
408 AspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysTh 424
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2118 CACTGGTATTGTGGAGTCCA..... 2138
424 rHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArg 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2139 .....GCATTTTGGCGTGA..... 2156
441 lyAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGly 457
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2157 .....CCCAACATGATCATCAAGCGGATTCATTCGTTAAGCCAA... 2198
458 LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPh 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2199 .....ATGGCGATGCGAACGCTTCTATCCCTACC..... 2228
474 eProAspPheValTyrLysArgIle.....LysAlaA 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2229 .CCACACCGGTTTATACAGAGAAATGTCGCTCATCATGTTAAAGCTA 2277
485 rGsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGly 501
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

2278 ATACGAT...CCAACATCACTTTTGTCTCAAGCGGCTTATGACAAA 2324
502 ProValHisGluValMet.....ValProAlaLysPr 512
::: ||| ::::: |||::: |||::: |||
2325 GGCATTAAAGAAAGAAATAGGACTTGAAGACAAAGTGTGGCGGTAA... 2372
512 oGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSerValP 529
::: |||
2373 .....AATGTC..... 2378
529 roProValArgAsnLeuHisGlnSerGlyPheSerLeuSerGlySerGln 545
|||||::: ::::: |||::: |||
2379 .....AGAAATATCACTTAAACACGCAATCAATTCACGACACTACT 2420
546 AlaAspAspHisIleAlaArgArgThr.....AlaGl 556
|||::: |||::: |||
2421 GCTCACATTGAAGTCAATCTGAACTTACCATGTGTCGTGGATGGCAA 2470
556 nLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeu 571
::: |||::: |||::: |||
2471 AGAAGTAACTTCTAAACACGCAATCAATGAAGTGGCGCAACTC 2516

seq_name: /cgn2_6/ptodata/1/1na/PCTUS_COMB.seq: PCT-US96-05800-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9605800
; GENERAL INFORMATION:
; APPLICANT: Oravax, Inc.
; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05800
; FILING DATE: 23-APR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,041
; FILING DATE: 28-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/568,122
; FILING DATE: 06-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T. 30,162
; REGISTRATION NUMBER: 06132/020001
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-05800-1

alignment_scores:
Quality: 136.50 Length: 632

```

```

Ratio: 0.493 Gaps: 29
Percent Similarity: 43.829 Percent Identity: 19.146

alignment_block:
US-09-367-496-8 x PCT-US96-05800-1 ..
Align seg 1/1 to: PCT-US96-05800-1 from: 1 to: 2735

17 LeuLeuIleArgGlyGlyArgIleValAsnAspGlnSerPheThrAl 33
|||||::: |||::: |||::: |||::: |||
1023 TTAATTATCACTAACGCTTTAATCGTGGATTACACCGGTATTATTAAGC 1072
33 aspValHisValGluAspGlyLeuIleLysGlnIleGlyGlu..... 47
|||||::: |||::: |||::: |||
1073 GGATTTGGTTATTAAAGATGGCAAAATCGCTGGCATTTGTTAAAGCGGTA 1122
48 .....AsnLeuIleValProGlyGly 54
|||||::: |||
1123 ACAAGACATGCAAGATGGGTTTAAACAACTCTAGCGTAGGTCTGCT 1172
55 IleHisThrIleAspAlaHisGlyLeuMetValIleuProGlyGlyValAs 71
|||||::: |||::: |||::: |||
1173 ACTGAAGCCTTAGCCGGTGAAGTTTGTATCGTAACGGCTGGTGTATTGA 1222
71 pValHisThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspA 88
|||||::: |||
1223 CACACATCCACTTTCATTTCCACC..... 1247
88 spPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly...ThrThrMet 103
|||||::: |||::: |||::: |||
1248 .....CAACAAATCCCTACAGCTTTTGCAGCGGTGTAAACAACCATG 1289
104 IleLeuAspHisValPheProAspThrGlyValSerLeuLeuAlaAlaTy 120
|||||::: |||
1290 ATTGGTGGTGAACCGGCTCTGCTGATGCACATAATGCG..... 1328
120 rGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAspTyrSerL 137
1328 ..... 1328
137 euHisValAspIleThrArgTrpHisGluSerIleLysGluLeuGlu 153
|||||::: |||::: |||::: |||
1329 .....ACTACTATCACTCCAGGCGAGAAATTTAAATGGATGCTCAGA 1373
154 AlaLeuValLysGluLysGlyValAsnSerPheLeuValPheMetAlaTy 170
|||||::: |||
1374 GCG...GCTGAAGAATATCTATGAAT..... 1397
170 rLysAspArgCysGlnCysSerAspSerGlnMetTyrGluIlePheSerI 187
1397 ..... 1397
187 leIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluAsnGlyAsp 203
|||||::: |||::: |||::: |||
1398 .....TTAGGTTTCTGGCTAAAGGTAAAGTCTTAAACGATCGC 1436
204 IleValGluGluGluGlnLysArgLeuLeuGlyIleThrGly... 219
|||||::: |||
1437 ACCTAGCCGATCAA.....ATTGAAGCGGTCGATGGCTT 1474
220 .....ProGlu.....GlyHisValL 225
|||||::: |||
1475 TAAATTCACGAAGACTGGGCGACCTCTCTGCAATCAATCATCGCT 1524
225 euSerHisProGluGluValGluAlaGluAlaValTyrArgAlaValThr 241
|||||::: |||::: |||::: |||
1525 TAGATGTTCCGACAAATACGATGTGCAAGTGCCTATCCACACACACT 1574
242 IleAlaLysGlnAlaAsnCysProLeuTyrValThrLysValMetSerly 258
::: |||::: |||
1575 TTG...AATGAAGCGGTTGT..... 1592
258 sGlyAlaAlaAspAlaIleAlaGlnAlaLysArgArgGlyValValValP 275

```

```

1593   ...GTAGAAGACATATGCTCTATTGCTGGACGCACTATGCACTT 1638
275   heGlyGluProIleThrAlaSerLeuGlyThrAspGlySerHisTyrTrp 291
1639 TC.....CACATGAGGCGTGGGGGGGACAC.....1667
292 SerLysAsnTrpAlaLysAlaAlaPheValThrSerProProValAs 308
1668 GCTCTGTATATTATTAAGTACCGGTGAACACAACTTCTCCGGCTTC 1717
308 nProAspPro.....ThrThrAlaAspHisLeuT 318
1718 CACTAACCCCACTCCCTTACCGTGAATACAGACGACGACATGG 1767
318 hrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCys 334
1768 ACATGCTTATGCTGTGCCACCACTTGGATAAAGCACTTAAGAAGATGTT 1817
335 ThrPheThrThrAla.....GlnLysAlaValGlyLysAspAs 347
1818 CAGTTCGCTGATTCAAGATCCGCCCTCAAAACCACTTCCGCTGAAGACAC 1867
347 n.....PheAlaLeuIleProGluGlyThrAsnGly. 357
1868 TTTGCATGACATGGGATTTTCAATCACCAGTTCTGACTCTCAAGCGA 1917
358 .....IleGluGluArgMetSerMetValTrpGluLysCysValAla 371
1918 TGGCCCGTGTGGTGAAGTTTACTAGAACTTGGCAACAGCTGACAAA 1967
372 Ser.....GlyLysMetAspGluAsnGlu.....379
1968 AACAGAAGAAGAAATTTGGCGCTTGAAGAAGAAAGGCGATGAACGACAA 2017
380 .....PheValAlaValThrSerThrAsnAlaAlaLysI 391
2018 CTTCCAGATCAACCGCTACTTCTCTAAATACACCACTTAACCCAGCGATCG 2067
391 lePheAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAla 407
2068 CTCATGGGATTAACGAGTATGAGTTCTAGTAGAAGTGGGCAAGTGGCT 2117
408 AspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysTh 424
2118 GACTTGGTATGTGGAGTCCA.....2138
424 rHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValGluCysArg 441
2139 .....GCATCTTTGGCGTGA.....2156
441 lYAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGly 457
2157 .....CCCAACATGATCATCAAGGCGGATTCATGCTTAAGCCAA...2198
458 LysMetPheValThrProGlyAlaGlyArgPheValProArgLysThrph 474
2199 .....ATGGCGATGCGACGCTTCTATCCCTACC.....2228
474 eProAspPheValTyrLysArgle.....LysAlaAla 485
2229 .CCACACCGGTTTATACAGAGAAATGTTGCTCATCATGCTAAGACTA 2277
485 rgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGly 501
2278 AATACGAT...GCCAACATCACTTTTGTCTCTCAAGCGGCTTATGACAAA 2324
502 ProValHisGluValMet.....ValProAlaLysPr 512
2325 GGCATTAAAGAAATAGGACTTGAAGACAAAGTGTGCCGGTAA...2372
512 oGlySerGlyAlaProAlaArgAlaSerCysProGlyLysIleSerValp 529

```

```

2373 .....AATTGC.....2378
529 roProValArgAsnLeuHisGlnSerGlyPheSerLeuSerGlySerGln 545
2379 .....AGAAATATCACTAAATAAAGACATGCAATTCACACGACACT 2420
546 AlaAspAspHisIleAlaArgArgThr.....AlaG1 556
2421 GCTCACATTTGAAGTCAATCTCTGAACCTTACCATGCTTCTGTTGGATGGCAA 2470
556 nLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeu 571
2471 AGAAGTAACCTCTAAACCCCACTAAAGTAGCTTGGGCGCAACTC 2516
seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-036-987A-1

seq_documentation_block:
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036.987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

```

alignment_scores:

```

Quality: 122.50      Length: 529
Ratio: 0.513        Gaps: 27
Percent Similarity: 45.180      Percent Identity: 21.739

```

alignment_block:

```

US-09-367-496-8 x US-09-036-987A-1 ..

```

```

Align seg 1/1 to: US-09-036-987A-1 from: 1 to: 80161

```

345 LysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGluGluArg 361
 36853CCCTGGCGGTTCCTGGCGCGGTCTCGACGGCGCTCGGCGCCA 36896
 361 GMetSerMetValTrpGluLysCysValAlaSer.....GlyLysMetA 376
 36897 GCGCGACGGCTGCGCGAGCTGCGAGCAGCACCGATCCCGACCGCGTCG 36946
 376 spGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePhe 392
 36947 ATGTCCGGTTCTACTGCGCCCGCACGGCGCAG..... 36978
 393 AsnPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAspLe 409
 36978 36978
 409 uValIleTrpAsnProLysAlaThrLysIleSerAlaLysThrHisA 426
 36979 ...GCTTGGGAGACCGAGCGGTG...GTCTTGGTCGGGACGCGCTA 37021
 426 snLeuAsnValGluTyrAsnIlePheGluGluValGluCysArgGlyAla 442
 37022 CGTTCGGCTCCGGCGTT.....GGCGTGTGTTCGCCGCGTGAA 37059
 443 ProAlaValIleSerGlnGlyArgValAlaLeuGluAspGlyLysMe 459
 37060 CCAGCGGTGATGCTCTGAGGGGAGCGTC...CTGGACGGCAGGTCGT 37106
 459 tPheValThrProGly.....AlaGlyArgPheVal..Pr 470
 37107 CTTCGTCTCCCGGTGAGGCGTGGCAGTGGCGGTATGCGAGTCGAC 37156
 470 oArgLysThrPheProAspPheValTyrLysArgIleLysAlaArgAsnA 487
 37157 TCCTGGAGCTTCGCCGAGCTTCGCCGCCACATGGACGAGTCGCCAC 37206
 487 tGLeuAlaGluIleHisGlyValProArgGlyLeuTyrAspGlyProVal 503
 37207 GCGTCGCGAGGTACGTGCTGCTGTCAGCTGCTGCGCGGAC 37256
 504 HisGluValMetValProAlaLysProGlySerGlyAlaProAlaArgAl 520
 37257 GGAGAA...CTCCCCACCGCTGGACCGGTGGAGTCTCCAGCGCGCT 37303
 520 aserCysProGlyLysIleSerValProValArgAsnLeuHisGlns 537
 37304 C..... 37304
 537 erGlyPheSerLeuSerGlySerGlnAlaAspAspHisIleAlaArgArg 553
 37305CTTCGCGT.....GATGGTGTGCTCGCTCGCGCA 37331
 554 ThrAlaGlnLysIleMetAlaProProGlyGlyArg 565
 37332 GGTGGCGGTCTCTACGGGTGAGCGCGCGCGCT 37367
 seq_name: /cn2_6/ptodata/1/ina/backfiles1.seq:5206163-2
 seq_documentation_block:
 Patent No. 5206163
 APPLICANT: RENARD, ANDRE;DINA, DINO;MARTIAL, JOSEPH
 TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
 VIRUS PROTEIN
 NUMBER OF SEQUENCES: 3
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/550,816
 FILING DATE: 06-JUL-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 331,037
 FILING DATE: 29-MAR-1989
 APPLICATION NUMBER: 752,981
 FILING DATE: 08-JUL-1985
 SEQ ID NO:2

5206163-2
LENGTH:12492

alignment_scores:

Quality: 114.50 Length: 360
Ratio: 0.748 Gaps: 13
Percent Similarity: 42.500 Percent Identity: 20.556

alignment_block:

US-09-367-496-8 x 5206163-2

Align seg 1/1 to: 5206163-2 from: 1 to: 12492

```

125 GluArgAlaAspSerAlaAaCysCysAspTyrSerLeuHisValAspIle 141
10678 GAAATTCAGCAGCCAGGATTCATGATAAATGTTAGAGATCTTTCA 10727
141 eThrArgTrpHisGluSerIleLys.....G 150
10728 CACAATAGCCCAACCCAGCCTAAGCACACTTACGCGAAGTGACGTGG 10777
150 LuGluLeuGluAlaLeuValLysGluLysGluValAsnSerPheLeuVal 166
10778 AACAGCTTGAGGAGGATCAACAGAAAGGGCTGCAGGCTTTCTA... 10824
167 PheMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrG 183
10825 .....GAAAAGAGAATCTTGGAGAGTACTGGGACTCAGAGAACCT 10868
183 uIlePheSerIleIleArgAspLeuGluAlaLeuAlaGlnValHisAla. 199
10869 GGTGGACCACTAATCAGAGACCTGAAACACAGGAGGAGATAGATATT 10918
199 ..... 199
10919 ATGAGACACCAATACCTAAGAACAGAGAGAGGATGTCAGTGAGATTGG 10968
200 GluAsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuG 216
10969 CAAGCAGGGGACATAGTTGATGAAGAAACCAAGA..... 11004
216 yIleThrGlyProGluGlyHisValLeuSerHisProGluLeuValGlu 233
11005 .....GTGATCAATACCTGAA..... 11022
233 laGluAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysPro 249
11023 .....GCTAAGACAGA 11034
250 LeuTyrValThrLysValMet..... 256
11035 CTGGCCATCACTAAAGTTATGTACAACTGGGTGACGACGACCTGTTGT 11084
257 .....SerLysGlyAlaAlaAspAlaIleAlaGlnAlaL 268
11085 GATCCAGGATGAAGGAGGAGGAGGATTTTCAAGATCTTTACAGAG 11134
268 ysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuGly 284
11135 TAAGAAAGGAATGGGACCTGTTCAATGAGCCAGTACCTGTGAGTTTGT 11184
285 ThrAspGly..... 287
11185 ACTAAGGCTGGGACCCAGCAGTACTAGTAGGATCTACGGCTTATTGG 11234
288 .....SerHisTyrTrpSerLysAsnTyrAlaLysAlaAlaLap 301
11235 TGAATTCAAAATATTACTACAGAGGAGTGCACAAATTCATCGATA 11284
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
11285 CCATCAGGACCATGTTGGAGTACCTAGCTATCATACAGCAGAT..... 11328

```

```

318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
11329 .....GTTGAAGTATACATA..... 11343
334 sThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 351
11344 .....AGAAATGGACAAGGGTAGTGGCCAGCAGAC..... 11376
351 leProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGlu 367
11377 ..ACAAGCGCAGGTAAACAGCATGCTAAATGTGTTAAACAATGATGTC 11424
368 LysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
11425 TTCGTGAAAGTACGGGGTT..... 11445
384 rSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 401
11446 .....CCATATAAGAGTTTAAAT.....AGAGTTGCAAGGA 11476
401 alaAlaVal...GlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
11477 TCATGCTGTGGGATGACGCTTCCTGATAACAGAGAGGGGCTGGGC 11526
417 ThrLysIleSerAlaLysThrHisAsn 426
11527 ACTAAATTTGCCAACAAAGGATGCAAC 11556

```

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-09-320-878-19

seq_documentation_block:

```

; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT FILING DATE: 1999-05-27
; EARLIER FILING DATE: 1999-05-27
; EARLIER FILING DATE: 1998-08-28
; EARLIER FILING DATE: 1998-08-28
; EARLIER FILING DATE: 1998-05-06
; EARLIER FILING DATE: 1997-04-30
; EARLIER FILING DATE: 1997-04-30
; EARLIER FILING DATE: 1999-02-08
; EARLIER FILING DATE: 1999-02-08
; EARLIER FILING DATE: 1998-09-22
; EARLIER FILING DATE: 1998-09-22
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

alignment_scores:

Quality: 112.50 Length: 478
Ratio: 0.511 Gaps: 26
Percent Similarity: 46.025 Percent Identity: 23.640

alignment_block:

US-09-367-496-8 x US-09-320-878-19

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

```

Align seg 1/1 to: US-09-320-878-19 from: 1 to: 38506
178 AspSerGlnMetTyrGluLeuPheSerIleIleArgAspLeuGlyAlaLe 194
||||| : : : : : |||||||
17421 GACTCGCGCTCGCGGGCTGTCGCGGAACCTCGCGGACCTGGCGCGAC 17470
||||| : : : : : |||||||
194 uAlaGlnValHisAlaGluAsnGlyAspLeuValGluGluGlnLysA 211
||| ||| : : : : : |||
17471 GGCACCGTC.....GTGACCTCGACCTACAGGACGCGGAGCGCGCG 17514
||| ||| : : : : : |||
211 rglLeuLeuGluLeuGlyIleThrGly.....ProGluGlyHisValLeu 225
||| : : : : : |||
17515 CCCGGCTGCTCGCGGGCTCTCCGACGCGACCGCGCTCAGCGCGCTCT 17564
||| : : : : : |||
226 SerHisProGluGluVal.....GluAlaGln 234
||| ||| : : : : : |||
17565 CACCTGCGCGCCACCGCTCGACTCGAGCGCGCTCGCGGACCGCGGA 17614
||| : : : : : |||
234 uAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeuT 251
||||| : : : : : |||
17615 CGCGCTCGCGCTGTCGTGAC.....GCGAGGCGCCACCGCGCGCTCC 17658
||| : : : : : |||
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
||||| : : : : : |||
17659 ACCTGACCGCTCTCTG.....CGGAGCGCGCGCTGCGCGGA 17696
||| : : : : : |||
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuGln 284
||| : : : : : |||
17697 GCGCGTCCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17746
||| : : : : : |||
284 yThrAspGlySerHisTyrTyrSerLysAsnTyrAlaLysAlaAlaLap 301
||||| : : : : : |||
17747 CGCGCGCGGTACGGCGCGCTACGCGCGGTACGCGCTCTCTCTCTCTCT 17796
||| : : : : : |||
301 he.....ValThrSerProVal 307
||| : : : : : |||
17797 TCGCGGTACAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17846
||| : : : : : |||
308 AsnPro..... 309
17847 AGCCCTTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17896
||| : : : : : |||
310 .....AspProThrThrAlaAspHisLeuT 318
||||| : : : : : |||
17897 CGCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17940
||| : : : : : |||
318 hrCysLeu.....LeuSerSerGlyAspLeuGlnValThr..... 329
||||| : : : : : |||
17941 CCGCGCTGGACACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17990
||| : : : : : |||
330 .....GlySerAlaHisCysThrPheThrThrAlaGlnLysAl 342
||||| : : : : : |||
17991 GACGTGCGACTGGTGGAGCTTCGCGCGCGCGCGCGCGCGCGCGCG 18040
||| : : : : : |||
342 aValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleG 359
||||| : : : : : |||
18041 CACC.....CTCTCGCGATCTCGCGCGCGCGCGCGCGCGCGCGCG 18081
||| : : : : : |||
359 luLuArgMetSerMetValTyrGluLysCysValAlaSerGlyLysMet 375
||||| : : : : : |||
18082 ACAGCAGCAGTCGACGCG.....GCGCGCGCGCGCGCGCGCGCGCG 18119
||| : : : : : |||
376 AspGluAsnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePh 392
||||| : : : : : |||
18120 CTGAGCGCGGAGCTCGGTGCTCTACCGCGCGCGGAA..... 18155
||| : : : : : |||
392 eAsnPheTyrProArgLysGlyArgValAlaValGlySerAspAlaAspL 409
||||| : : : : : |||
18156 .....CAGCAGCGCGGTATCAGGAGTTGTCGCGCGCGCGCGCGCG 18198
||| : : : : : |||
409 euValleTyrAsnProLysAlaThrLysIleSerAlaLysThrHis 425
||||| : : : : : |||
18199 TGGTCTCAACCAACCCCTCCCGCGAGGCGGTGCGACACGCGCGCGCG 18248
||| : : : : : |||

```

seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-000-016-1

seq_documentation_block:

; Sequence 1, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

; APPLICANT: Akira ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE

; TITLE OF INVENTION: ITS EXPRESSION PRODUCT

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/000,016

; FILING DATE: January 30, 1998

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:


```

698 CGAGTTCGGTGATGTCGAAGACGCCCGCGTGGACCTTGGCGG..... 658
699 CITY: Indianapolis
700 STATE: Indiana
701 COUNTRY: USA
702 ZIP: 46268
703 COMPUTER READABLE FORM:
704 MEDIUM TYPE: Floppy disk
705 COMPUTER: IBM PC compatible
706 OPERATING SYSTEM: PC-DOS/MS-DOS
707 SOFTWARE: Patentin Release #1.0, Version #1.30
708 CURRENT APPLICATION DATA:
709 APPLICATION NUMBER: US/09/036,987A
710 FILING DATE: 09-MAR-1998
711 CLASSIFICATION: 435
712 ATTORNEY/AGENT INFORMATION:
713 NAME: Stuart, Donald R
714 REGISTRATION NUMBER: 28,479
715 REFERENCE/DOCKET NUMBER: 50,608
716 TELECOMMUNICATION INFORMATION:
717 TELEPHONE: (317)337-4816
718 TELEFAX: (317)337-4847
719 INFORMATION FOR SEQ ID NO: 1:
720 SEQUENCE CHARACTERISTICS:
721 LENGTH: 80161 base pairs
722 TYPE: nucleic acid
723 STRANDEDNESS: double
724 TOPOLOGY: linear
725 MOLECULE TYPE: DNA (genomic)
726 US-09-036-987A-1
727
728 alignment_scores:
729 Quality: 110.00 Length: 685
730 Ratio: 0.364 Gaps: 35
731 Percent Similarity: 44.088 Percent Identity: 20.146
732
733 alignment_block:
734 US-09-367-496-8 x US-09-036-987A-1/rev ..
735
736 Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161
737
738 21 GlyGlyArgIleValAsnAspGlnSerPheTyrAlaAspValHisVa 37
739 9704 GCGGAATCGTGTGAACGGGTGCAACAGACAGCACCGTCCGGTGTCCATGTT 9655
740 37 lgluaspGlyLeuIleLysGlnIleGlyGluAsnLeuIleValProGlyG 54
741 9654 TGACCATGTGTTCCTGAC.....GAAGGAGAGCCAGCACCTGGCC 9615
742 54 lylleHisThrIleAspAlaHisGlyLeuMetValLeuProGlyGlyVal 70
743 9614 GAGCGTCAACGACGAGAGTGTGCGAGGACTGGCACCTTCTCTCGACTT 9565
744 71 .....AspValHI 73
745 9564 CCACCACCATGTTCGGTTCGGTTCGGTAGAACGAGACCCCTGTTC 9515
746 73 sThrArgLeuGlnMetProValLeuGlyMetThrProAlaAspPheC 90
747 9514 GACTGGAGCACGCTTACACGACGCGCGCGCGCCGCGCC..... 9475
748 90 ysgInGlyThrLysAlaAlaLeuAla.GlyGlyThrThrMetIleLeuAs 106
749 9474 ..CGAGGAGTAGTTCACATAGGCGGGAACGCTGCCACGCTGTGCTGGC 9427
750 106 pHisValPheProAspThrGlyValSerLeuLeuAlaAlaTyrGluGlnT 123
751 9426 TGATGTGCTCCG.....AGTTGCTCGAACCGTC 9398
752 123 rfArgGluArgAlaAspSerAlaAlaCysCysAspTyrSerLeuHisVal 139
753 9397 GCGGAGAGCTCGAGGAC.....GTT 9378
754
755 seq_name: /cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-036-987A-1
756
757 seq_documentation_block:
758 ; Sequence 1, Application US/09036987A
759 ; Patent No. 6143526
760 ; GENERAL INFORMATION:
761 ; APPLICANT: Baltz, Richard H.
762 ; APPLICANT: Broughton, Mary C.
763 ; APPLICANT: Crawford, Kathryn P.
764 ; APPLICANT: Madduri, Krishnamurthy
765 ; APPLICANT: Merlo, Donald J.
766 ; APPLICANT: Treadway, Patti J.
767 ; APPLICANT: Turner, Jan R.
768 ; APPLICANT: Waldron, Clive
769 ; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
770 ; TITLE OF INVENTION: Production
771 ; NUMBER OF SEQUENCES: 39
772 ; CORRESPONDENCE ADDRESS:
773 ; ADDRESSEE: Dow Agrosciences LLC Patent Department

```

```

8634 ..... |||||:..: |||||:..: CCTCGTGGAAATCGGTTGGAGC 8612
403 1..... GlySerAspA 407
      :..... |||||:..:
8611 GCNAACCTCTGGGAGATCAACACCCCATTTGGACGATTCTGGACAAGAAAT 8562
407 laAspLeuValIleTetPAsnProLysAlaThrLysIleIleSerAlaLys 423
      :|||:..:
8561 CTGACTTC.....GGACATCGCTGCGATCGTCCCGG 8530
424 ThrHisAsnLeuAsnValGluTyrAsnIlePheGluGlyValIleCysAr 440
8529 TCACGCGGGAATCCGAGTCGACGCAACGCGATGAACACCTGCCA..... 8485
440 gGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspG 457
      :|||:..:
8484 ....AACCCTGGCGTCCATGTTTC...GGCCACACGTTGTCACGGATG 8442
457 Ly.....LysMetPheValThr 462
      :|||:..:
8441 GCCATCAGACTGCCCATGTCGCGCGGTTCTCAGTTTCGCAACTAGCTTTT 8392
463 ProGlyAlaGlyArgPheValProArgLysThrPheProAspPheVal. 478
      :|||:..:
8391 CCAGAGTGGGAACGATCTGCTGCGGGTGGGACATTTCTCTGATTTCACGG 8342
479 ....TyrLysArgIleLysAlaArgAsnArgLeuAlaGlu..... 490
8341 CGGATCAGCGCGGTTGGCAGCAACGCGCTCGTCGACGCGCGGC 8292
491 .....IleHisGlyValProArgGlyLeuTyrAsp.. 500
8291 CAGCTGACCTCGCAGCTCGTCTTCGGTAAACGTCGCGGGTCTCGAGGACCA 8242
501 .....GlyProValHisGluValMetValProAlaLys.ProGlySer 514
8241 GACCGGCTCCCGATCGCTAGGAGCTCCGCCCTACGAGATTGCTCCAG 8192
515 GlyAlaProAlaArgAlaSerCysProGlyLysIleSerValProPro.. 530
8191 AAGTCCCGGGAGATCAACTCGCGGTACGCCGTTGACCGTGGCGGTTTC 8142
531 .....ValArgAsnLeuHisGlnSerGlyP 539
8141 CTGCGTGGTCTGAGCCCATGATGATGATCACTGAACACGACTCCAGCA 8092
539 heSerLeuSerGly..... 543
8091 GTTCGTTGACGGTACGTATTCTGTGACCCGGAGCTTCGAGGCGCACTCC 8042
544 .....SerGlnAlaAspAspHisIleAlaArgArgThrAlaG1 556
8041 CCCATCTCCCGTACTTCGTCTGACAGAAGGTGGCGATCACCTCGACGTC 7992
556 nLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeuSer 572
7991 GAGCTGCGCCGCCG.....CGCAACAACTTTCCACCAATGCT 7952

```

us-09-367-496-8.std.rni

Tue Jul 31 13:08:19 2001

Tue Jul 31 13:08:18 2001

OM of: US-09-367-496-8 to: N_Geneseq_0601.* out_format : pfs

Date: Jul 30, 2001 12:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09367496/runat_30072001_081221_27409/app_query.fasta_1.637
-DB=N_Geneseq_0601 -OFMT=fastq -SURFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09367496_CGNI_1_224 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-367-496-8

Query length: 572

Database: N_Geneseq_0601.*

Database sequences: 730101

Database length: 313950809

Search time (sec): 112.950000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60818		2879.00	4602.51	1690	2.2e-248
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60817		2803.00	4478.99	1920	1.6e-241
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60815		2325.00	3711.45	1817	9.3e-199
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60816		2151.00	3429.29	2297	4.8e-183
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60814		2044.00	1651.55	2052	5.0e-84
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60813		1018.00	1611.65	1734	8.5e-82
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60812		837.00	1319.57	1569	2.8e-81
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60811		837.00	1305.57	1938	1.6e-65
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60810		805.00	1076.22	1383	9.5e-52
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60809		698.50	1076.22	1385	5.6e-52
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60808		691.50	1089.50	1374	1.0e-52
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60807		684.50	1071.66	2518	1.0e-51
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60806		684.50	1069.09	3190	1.4e-51
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60805		600.00	933.82	3039	4.8e-44
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60804		415.00	662.96	288	5.9e-29
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60803		409.50	636.31	1377	1.8e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60802		408.50	634.70	1377	2.2e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60801		408.50	645.86	459	5.3e-28
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60800		407.50	633.10	1377	2.7e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60799		407.50	633.10	1377	2.7e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60798		407.50	633.10	1377	2.7e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60797		407.50	633.10	1377	2.7e-27
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60796		338.50	519.92	1701	5.5e-21
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60795		306.00	501.48	334	5.8e-20
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60794		309.50	441.12	32768	1.3e-16
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60793		252.00	382.55	1464	2.4e-13
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60792		241.50	343.39	11340	3.7e-11
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60791		240.00	361.04	1795	3.9e-12
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60790		239.00	345.54	6433	2.6e-11
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60789		238.00	340.22	9048	5.6e-11
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60788		225.50	339.29	1557	6.3e-11
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60787		205.00	320.00	444	7.5e-10
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60786		205.00	299.30	2057	1.1e-08
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60785		202.50	314.27	387	1.6e-09
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60784		200.50	295.35	2201	1.8e-08
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60783		197.50	270.24	1880	2.6e-08
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60782		178.50	292.02	876	4.5e-07
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60781		157.00	231.90	1216	6.0e-05

/SID88/gcgdata/geneseq/geneseq/NA1997.DAT:AAV74333 - 151.50 205.75 0.0017
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV93530 + 143.00 209.99 0.0010
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV93531 - 143.00 209.99 0.0010
/SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV93532 + 142.00 200.40 0.0034
/SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAV92097 + 142.00 192.93 0.0089

seq_name: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60818

seq_documentation_block:

ID AAV60818 standard; cDNA; 1690 BP.

AC AAV60818;

XX 08-DEC-1998 (first entry)

DT Human partial ULIP-4 coding sequence.

DE Human; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;

KW neurodegenerative disorder; diagnosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 32..1690

FT /tag= a

FT /product= "partial ULIP-4"

FT /note= "Unc-33-like phosphoprotein 4"

FT /transl_except= (pos: 197..199, aa: xaa)

FT /note= "Xaa = unknown; codon at this position is given in sequence as a stop codon and may alter the reading frame"

XX FR2759701-A1.

XX 21-AUG-1998.

XX 19-FEB-1997; 97FR-0001961.

XX 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

XX Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX P-PSDB; AAV68489.

XX Mouse and human ULIP poly:peptide(s) - useful in detection of

XX para-neoplastic neurological syndromes

XX Claim 3: Fig 12; 90pp; French.

XX This sequence represents a partial human Unc-33-like phospho-protein

XX (ULIP)-4 coding sequence. The sequence was isolated based on similarity

XX to the rat ULIP sequence. Proteins of the ULIP family or their

XX corresponding nucleic acids can be used in compositions for treating

XX neurodegenerative disorders and neoplasms, especially for para-neoplastic

XX neurological syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 1690 BP; 359 A; 505 C; 539 G; 287 T; 0 other;

alignment_scores:

Quality: 2879.00 Length: 553

Ratio: 5.216 Gaps: 0

Percent Similarity: 99.819 Percent Identity: 99.819

alignment_block:

US-09-367-496-8 x AAV60818 ..

Align seg 1/1 to: AAV60818 from: 1 to: 1690

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 32 ATGTCTCTCCAGGGCAAGAAAGCATCCCGGATCAGAGTGACCGCT 81
 17 UleulleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34
 82 TCTGATCAGAGTGGAGAGTCGTGAATGAGCAGCAGTCTTTACGGTG 131
 34 spValHisValGluAspGlyLeulleLysGlnIleGlyGluAsnLeu 50
 132 ATGTGCAGCTGGAAGATGCTTGTATAAAACAAATCGGAAACCTCATC 181
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 182 GTCCCTGGGGCATCTAGACCACTTGAGCCGACGCGCTGATGCTCTTC 231
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 232 TGGTGGGGTTCAGCTCCACAAAGGCTGCAGATGCCGTCTCTGGGATGA 281
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGly 100
 282 CACGGCTGACGACTCTGTCTAGGGGACCAAGCAGCGCTAGCAGGAGA 331
 101 ThrThreMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
 332 ACCACCATGATCTTGACCACTGCTTCCCGGACACGGGTGTGAGCTGCT 381
 117 uAlaAlaTyrGluGlnTyrArgGluArgAlaAspSerAlaAlaCys 134
 382 GCGGGCTACGAGCAGTGGCGGAGCGGCGACAGCGCGCTGCTGCG 431
 134 spTyrSerLeuHisValAspIleThrArgTyrPheHisGluSerIleLys 150
 432 ACTACTCCCTGACGTGGACATCACCCGATGCCATGAGAGCATCAAGAG 481
 151 GluLeuGluAlaLeuValLysLysGlyValAsnSerPheLeuValPh 167
 482 GAGCTGAGGGCTGCTCAAGGAGAGGGTGTGAACCTCTCTCTGCTT 531
 167 eMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
 532 CATGGCATACAGAGACCGGTGCCAGTGCAGCAGCAGCAGCATGACGAG 581
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 582 TCTTCAGCATCATCCGGACCTGGGGGCTTGGCCCGCAGGTGCACGTGAG 631
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyI 217
 632 AACGGGACATCTGGAGGAGGAGCAGACGCTTCTCTGAGCTCGGCAT 681
 217 eThrGlyProGlyHisValLeuSerHisProGluGluValGluAlaG 234
 682 CACTGGCCCGAGGGCCAGTGTCTAGCCACCCCGAGAGGTGGAGGTG 731
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 732 AGCGGTGTACGAGCTGCACCATCGCCCAACAGCAGCAACTGCCCGTG 781
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 782 TAGCTACCAAGGTATGAGCAAGGGGCGCGCAGCATCGCTCAGGC 831
 267 aLysArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
 832 CAAGCGCAGAGGGTGTCTGTTTGGGAGCCCATCACCGCCAGCGCTG 881
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 882 GCACCGAGGTTCACACTACTTGGAGCAAGAACTGGGCCAAGGCTGACGC 931
 301 PheValThrSerProProValAsnProAspProThrAlaAspHisLe 317

932 TTGTCACATCACCCCTGTCAACCCAGACCCACCACGCGCACCACT 981
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHis 334
 982 CACTCTGCTCTCTCCAGCGGGACCTTCAGGTGACAGGCGCGCACT 1031
 334 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1032 GCACCTTCACCACTGCCAGAGGCTGTGGCAAGGACAACCTTCGCGCTG 1081
 351 ileProGluGlyThrAsnGlyIleGluArgMetSerMetValTrpGl 367
 1082 ATCCCGAGGGCACCACCGCATTTGAGGAGCGCATGTGATGGTGGGA 1131
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
 1132 GAAATGTGTGGCTCTGGGAAGATGGAGAGAAATGAGTTCTGCGGGTGA 1181
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400
 1182 CCAGTACAAATGCTGCCAAATCTTCAATTTTACCCCAAGGAGGCGCA 1231
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
 1232 GTGCTGTGGGCTCTGACGTGACCTGTCTATATGAAACCCCAAGGCCAC 1281
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
 1282 CAAGATCATCTGCCAAGACCCCAATCTGAACGTGGAGTACAACATCT 1331
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 1332 TCGAGGAGTGGAGTGGCGGGAGCGCTGCCGTGTCATAGTCAGGGC 1381
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1382 CGAGTGGCGCTGGAGCAGGGAAGATGTTGTACCCCGGGGGCGCG 1431
 467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLys 484
 1432 CTTCGTCCCTCGGAAACATTCGCGACTTGTCTACAAGAGGATCAAG 1481
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
 1482 CTCGCAACAGGCTGGCGGAGATCCACGGTGTGCCCGCTGTATGAC 1531
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1532 GGGCCGTCTCCAGAGGTGATGTGCTGCCAAGCAGGAGTGGCGCTCC 1581
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534
 1582 GGGCGCGCGCTCTCCAGCAGCAAGATCTCGTGCTCTCTGTGGCAACC 1631
 534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspAspHisIle 550
 1632 TACATCAGTGGGGTTTCAAGCTATCTGGGTCTCAGGCTGATGACCATC 1681
 551 AlaArgArg 553
 1682 GCCGAGCGC 1690

seq_name: /SID8/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60817

seq_documentation_block:

ID AAV60817 standard; cDNA; 1920 BP.

XX AAV60817;

XX 08-DEC-1998 (first entry)

XX Mouse ULIP-4 coding sequence.

DE

us-09-367-496-8.std.rng

Tue Jul 31 13:08:18 2001

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
 KW neurodegenerative disorder; diagnosis; ss.
 XX Mus musculus.

Location/Qualifiers

102..1820

/*tag= a

/product= "ULIP-4"

/note= "Unc-33-like phosphoprotein 4"

FR2759701-A1.

21-AUG-1998.

19-FEB-1997; 97FR-0001961.

19-FEB-1997; 97FR-0001961.

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

XX Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX P-PSDB; AAM68488.

XX Mouse and human ULIP poly:peptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes

XX Claim 3; Fig 11; 90pp; French.

XX This sequence represents the mouse unc-33-like phospho-protein (ULIP)-4
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumourigenesis.

XX Sequence 1920 BP; 440 A; 524 C; 562 G; 394 T; 0 other;

alignment_scores:

Quality: 2803.00

Ratio: 4.988

Percent Similarity: 98.252

Percent Identity: 92.657

alignment_block:

US-09-367-496-8 x AAV60817 ..

Align seg 1/1 to: AAV60817 from: 1 to: 1920

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgIle 17

102 ATGTCCTTCACAGCAGAGAGCATTCCCGGATACAGCAGCAGCCCT 151

17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAla 34

152 TCTCATCAAGGCTGGAGAGATTGTGAACGATGACCATCTCTTCATGCTG 201

34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeu 50

202 ATCTGTATGTGAAGACCGCTGATTAACAAATTTGGAGAAATCTCATC 251

51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeu 67

252 GTCCCTGGGGGATCAAAACCATCATGCTCATGCGCTGATGGTGGCTGCC 301

67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84

302 TGGGGGAGTTGACGTTACACCCCGCTGCAGATGCTGTGATGGCATGA 351

84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 352 CCCAGCTGATGATTTCTGTCAGGCGCACCAAGCGGCTCTAGCAGCGGG 401
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
 402 ACCACCATGATATTGGACATGTGTTTCTGACGCTGTGTGAGCTGCT 451
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 452 GGCAGCTATGAGCAGTGGCGGAGGAGCAGCAGCGCGGCTCTGTGTG 501
 134 spTyrSerLeuHisValAspIleThrArgTyrHisGluSerIleLysGlu 150
 502 ACTACTCTTACATGTGGACATTCCTCGCTGGCAGCAGACCAAGAA 551
 151 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 552 GAGCTGGAGGCCCTAGTCAGGAGCAAGGTGTGAACCTCTCTGTGTCT 601
 167 eMetalatyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
 602 CATGSCATACAAAGGAGCAGGTGCCAGTGTACTGACGGTCAGATATATA 651
 184 lePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 652 TCTTCAGCCTCATCCGGACCTGGGAGCTGTGGCCAGGTGCACGAGAA 701
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGlyIle 217
 702 AATGGGAGCATCGTGGAGGAGGAACAGAGCGCTGTGGAGCAAGCAT 751
 217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
 752 CACTGGTCTCAGGCGCATGTGCTCAGCCACCCAGAGAGTAGAGCGCG 801
 234 luAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 802 AGGCTGTGTACAGAGCAGTCACCATTTGCCAGCAGCGCCACTGCCACTA 851
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGln 267
 852 TAGTTCACCAAGGTGATGAGCAAGGTGCAGCTGATGGTTGCCCAAGC 901
 267 aLysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
 902 CAA3CGCAGGGGGTGTCTCTTTGGGAACCTATCACTGCCAGCCTGG 951
 284 lyThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAla 300
 952 GCATGATGGCTCACACTACTGGAGCAAGAACTGGGCCAAGGCTGCAGCC 1001
 301 PheValThrSerProProValAsnProAspProThrThrAlaAspHisLe 317
 1002 TTTGTCACTTACCCCTATCAACCCGACCTTACTACTGCAGACCACT 1051
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1052 CACCTCTCTGCTGCTCCAGTGGGACCTCCAGGTGACAGGCGTGCCTC 1101
 334 ysThrPheThrThrAlaGlnLysAlaValAlaGlyLysAspAsnPheAlaLeu 350
 1102 GCNCCCTCACTACTGCCCGAGAGGCTGTGGCAAGACAACTTCCACTG 1151
 351 IleProGlyGlyIleThrAsnGlyIleGluGluArgMetSerMetValTrp 367
 1152 ATCCCGGAGGTAGTCAACGGTATAGAGAGCGCATGTCTGTGGTCTGGGA 1201
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
 1202 GAAATGTGTGGCTTCAGGAAATGGAGAGAAATGAGTTCGTCCTGCGG 1251
 384 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 400

```

|||||
1252 CCAGCACAATGCTGCCAAATCTTCAATTTTACCCAGGAAGGGCGT 1301
401 ValAlaValGlySerAspAlaAspLeuValIleTTPAsnProLysAlaTh 417
|||||
1302 GTGGCGGTGGCTCTGATGCTGACCTGCTATCTGGAACCCAGGGCCAC 1351
417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIleP 434
|||||
1352 GAAAGTCATCTCTGCAAGAGCCATAACCTGAATGTAGAGTACAACATCT 1401
434 heGluGlyValIleCysArgGlyAlaProAlaValValIleSerGlnGly 450
|||||
1402 TTGAAGGAGTGGAGTGGCGAGGAGTGGCCACGGTGGTCAATGTCAGGGC 1451
451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
|||||
1452 AGAGTGGTGTGGAGGACGAAACCTGCTGTCTCACTCCAGGGCTGGCCG 1501
467 gPheValProArgLysThrPheProAspPheValTyrLysArgIleLysA 484
|||||
1502 CTTTCATCCCGGAGAGCTGTCGGGACTTGTCTATAGAGGATAAAGG 1551
484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyrAsp 500
|||||
1552 CTCGCAACAGGCTAGCAGAGATCCACGGTGTGCTCGAGGCTGTACGAC 1601
501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
|||||
1602 GGGCTGTGCATGATGATGTTACTGCGAAGCCAGGAAGTGGCACACA 1651
517 oAlaArgAlaSerCysProGlyLysIleSerValProValArgAsnL 534
|||||
1652 GCGCGTGCATCTGTCAGGCAAGATCTCAGTCCACCCCGTGGCAACC 1701
534 euHisGlnSerGlyPheSerLeuSerGlySerGlnAlaAspHisIle 550
|||||
1702 TGCACAGTGGGGTTCAGCTACTGTGCTCTCAGGCTGAGATCACATT 1751
551 AlaArgArgThrAlaGlnLysIleMetAlaProProGlyGlyArgSerAs 567
|||||
1752 GCCAGACGTACGGCTCAGAGATCATGCGACCCCGGAGGACGCTCCA 1801
567 nIleThrSerLeuSer 572
|||||
1802 CATCAGCTCTCTTCC 1817

```

seq_name: /STDSB/gcgdata/geneseq/geneseq/NA1998.DAT:AAV60815

seq_documentation_block:

ID AAV60815 standard; cdna; 1817 BP.

XX AC AAV60815;

XX DT 08-DEC-1998 (first entry)

XX DE Mouse ULIP-1 coding sequence.

XX KW Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;

XX KW neurodegenerative disorder; diagnosis; ss.

XX OS Mus musculus.

XX FH key Location/Qualifiers

XX CDS 23..1741

XX FT /*tag= a

XX FT /product= "ULIP-1"

XX FT /note= "Unc-33-like phosphoprotein 1"

XX FR2759701-AL.

XX PD 21-AUG-1998.

XX XX

PF 19-FEB-1997; 97FR-0001961.
 XX
 PR 19-FEB-1997; 97FR-0001961.
 XX
 PA (INRM) INSM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
 PI Quach TT, Sobel A;
 XX
 DR WPI: 1998-449610/39.
 DR P-PSDB; AAW68486.
 XX
 PT Mouse and human ULIP polypeptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes
 PS
 PS Claim 3; Fig 9; 90pp; French.

CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

SQ Sequence 1817 BP; 426 A; 506 C; 511 G; 374 T; 0 other;

alignment_scores:

Quality: 2325.00

Ratio: 4.429

Percent Similarity: 91.944

Percent Identity: 75.306

alignment_block:

US-09-367-496-8 x AAV60815

Align seg 1/1 to: AAV60815 from: 1 to: 1817

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 |||||
 23 ATGCTCTATCAGGGGAGAAATAATATTCACCCCATCAGGAGGATCGTCT 72
 17 uleulleArgGlyGlyArgIleValAsnAspAspGlnSerPheTyrAlaA 34
 |||||
 73 TCTGATCAAAAGGTGGCAAGATGTGAATGATGACCATCTCTTCTATGAG 122
 34 spValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIle 50
 |||||
 123 ACATATACATGGAAGATGGTGTGATCAAGCAAAATAGGAGAAACCTGAT 172
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 |||||
 173 GTACCAGAGGGGTGAAGACCATCGAAGCCACCTCCAGATGGTGATTC 222
 67 oGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 |||||
 223 CGGAGGAATTCACCTGCATCTACTCGCTTCCAGATGCCTGACCAAGGAATGA 272
 84 hrProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 |||||
 273 CATCCGCTGATGACTTCTTCAGGGAACCAAGCGGCCCTGGCGGGGGA 322
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 |||||
 323 ACCACCATGATCATTCACCATGTTGTTCTTCCAGCCCGGAGGAGGCTATT 372
 117 uAlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysA 134
 |||||
 373 GGCTGCTTGTATCAGTGGAGGAGTGGGCTGACAGCAAGTCTGCTGTG 422
 134 spTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 |||||
 423 ACTATTTCGCTGCACGTGGACATCATGTAGTGGCAACAGGGCATCCAGGAG 472

11 GluLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
12 GAGATGGAAGCTCTGTTGAAGGACACCGGGGTAAATCTCTTCTCGTGA 522
13 eMetAlaTyrllysAspArgCysGlnCysSerAspSerGlnMetTyrlGluI 184
14 CATGGCTTCAAGATCATGATCCAGCTGACGGATTCACAGATCATGAAG 572
15 lePheSerIleIleArgAspLeuGlyAlaLeuGlnValHisAlaGlu 200
16 TGCTGAGCGTATCGGGATATCGGTGCCATAGCTCAAGTCCACGCGAG 622
17 AsnGlyAspIleValGluGluGluGlnLysArgLeuLeuGlyIle 217
18 AATGGTGACATCATGCTGAGGCACAGCAGAGGATCTGGATCTGGGCAT 672
19 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAla 234
20 CACGGGCCCGAGGAGACAGTGTGAGCGGCCACAGAGGAGTCCAGCGTG 722
21 luAlaValTyrlArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
22 AAGCTGTGAACCGGTCCATCATCTATTGCGCAACAGACCACTGCCCTCTG 772
23 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
24 aLysArgGlyValValValPheGlyGluProIleThrAlaSerLeuG 284
25 ACGGAAGAAGGAACTGTGTGTATGCTGAGCCCATCCAGCCAGCGCTGTG 872
26 lyThrAspGlySerHisTyrlTrpSerLysAsnTrpAlaLysAlaAla 300
27 GGACTGATGGCTCTCATCTACTTGGAGCAAGAACTGGGCCAAGGCTGGCC 922
28 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
29 TTTGTACCTCCCAACCTTGGAGCCCGGACCCCAACCACTCCAGACTTCT 972
30 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHis 334
31 CAACCTCGTGTCTCTGTGGAGACCTCCAGTCACTGGCAGTGCCCACT 1022
32 yThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
33 GCACCTTCAACACTGCCCAAGGCTGTGGGAGGAGCACTTCCACTTG 1072
34 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
35 ATTCCCGAGGGCCCAACAGGCCTCAGGAGCGGATGTGTGTCATTTGGGA 1122
36 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
37 TAAAGCTGTGTCTCACTGGGAAGATGGATGAGATCAGTTTGTGGCTGGA 1172
38 hrSerThrAsnAlaAlaLysIlePheAsnPheTyrlProArgLysGlyArg 400
39 CCAGCACCAACGAGCCCAAGTCTTCAACCTTACCCCGGAAAGAGTCGC 1222
40 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaTh 417
41 ATCTCGTGGGATCTGATGTGATCTGGTCACTTGGGATCTGGGACCTGACAGTGT 1272
42 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrlAsnIleP 434
43 GAAGACCACTCTGCCAAGACACACACACACTGCTCTTGAGTACACATCT 1322
44 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
45 TTGAAGGATGAGTGTGCGGCTCCCACTGGTGTGTCATGCCAGGCG 1372
46

XX

CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3
 CC coding sequence. The sequence was isolated based on similarity to the
 CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
 CC nucleic acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 2297 BP; 511 A; 653 C; 632 G; 501 T; 0 other;
 SQ

alignment_scores:

Quality: 2151.00 Length: 571
 Ratio: 4.226 Gaps: 0
 Percent Similarity: 89.142 Percent Identity: 69.002

alignment_block:

US-09-367-496-8 x AAV60816 ..

Align seg 1/1 to: AAV60816 from: 1 to: 2297

1 MetSerPheGlnGlyLysLysSerIleProArgIleThrSerAspArgLe 17
 270 ATGTCCTATCAGGGAAGAAGACATCCCGACATCACCAGTGACCGGCT 319
 17 uLeuIleArgGlyGlyArgIleValAsnAspAspGlnSerPheThrAla 34
 320 CCTCATCAGAGTGGACGCATCATCATGATGACCATCTCTTCTACGGCG 369
 34 spValHisValGluAspGlyLeuIleGlyGlnIleGlyGluAsnLeuIle 50
 370 ATGCTACCTAGAGTAGGAGTACATATAAACAATAGGAGAGAACCTGAT 419
 51 ValProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPr 67
 420 GTTCTGTGGAGTGAAGACCATCAGGCGAAGGCCGAATGGTCTATCC 469
 67 OGlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMet 84
 470 CGTGGCATTGATGTCACACTTACCTGACAGAGCCCTCCCGAGGCATGA 519
 84 hrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGly 100
 520 COTCGCTGATGACTTCTTCAGGCGACTAAAGCAGCGCTGGCAGGTGA 569
 101 ThrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLe 117
 570 ACCAGATGATCATGACCGTGTGTTCTCAACCTGGTCCAGCTGTGTT 619
 117 uAlaAlaTyroGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCys 134
 620 GACTTCCTTTGAGAAATGGCAGCAGCAGCAGCAGCAGCAGCAGCAG 669
 134 spTyroSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 150
 670 ACTATTCCTCCAGTGGACATCAGACGCTGTATGATGTTTCGGAA 719
 151 GlutGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPh 167
 720 GAGTGGAGTGTGTGGAGGACAAAGGTGTCACTCTTCCAGTCTA 769
 167 eMetAlaTyroLysAspArgCysGlnCysSerAspSerGlnMetTyroGlu 184
 770 CATGGGTATAGGACCTCTACCATGATGTCGACAGCCAGCTGTATGA 819
 184 IePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 200
 820 CCTTACCTCTCTTAAGGTTTGGAGCTGTGATCTTGTAGTCCATGCA 869
 201 AsnGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
 870 AATGAGATTGTAGTAGTCAGAGCAAAAACGGATCCTGGAGATGGCAT 919

217 eThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
 920 CACGGGTCCTCCAGGCTCATGCTCTGACGACACCCAGGAGCTGGAGCCG 969
 234 luAlaValTyroAlaValThrIleAlaLysGlnAlaAsnCysProLeu 250
 970 AGCTGTGTTCCGGGCTATTGCCATTGACGGCGGATCAATTGCCCTGTG 1019
 251 TyrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
 1020 TACATCACCAGGTCTATGACGAGAGTGCAGCGACATCATCGCATGGC 1069
 267 aLysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
 1070 CAGGAAGAAGGCCCTCTGTCTTCGTTGAGCCATAGCCGACGCTGG 1119
 284 LysThrAspGlySerHisTyroTrpSerLysAsnTrpAlaLysAlaAla 300
 1120 GAACCGATGGCACCCACTACTGGAGCAAGACTGGGCCAAGCAGCTGCA 1169
 301 PheValThrSerProValAsnProAspProThrThrAlaAspHisLe 317
 1170 TTTGTGACTTCCCTCCCTGAGCCAGACCCACCACCTCCTGACTACTT 1219
 317 uThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
 1220 GACCTCTTGTGCTGTGCTGTGAGACTTTCAGGTACAGGTAGTGGCCACT 1269
 334 ysThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 350
 1270 GTCCCTACAGTATTGCTCAGAAAGCTGTGGCAAGAGCAACTTCACTCTG 1319
 351 IleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpG 367
 1320 ATCCCTGAGGGTGTCAATGTTATAGAAGCGGTGACCGTTGTCTGGGA 1369
 367 uLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVal 384
 1370 CAAGGCAGTGGCTACTGGCAAGATGGATGAGAACCAAGTGTAGCCGTCA 1419
 384 hrSerThrAsnAlaAlaLysIlePheAsnPhetTyroArgLysGlyArg 400
 1420 CCAGCACCACCAAGCAGCAAGATCTTCAACCTGTATCCCGAGGAAAGTGG 1469
 401 ValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaPh 417
 1470 ATCGCTGTGGCTCCGATCTGCTGAGTGTCTTGGGACCCAGATAAAGT 1519
 417 rLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyroAsnIleP 434
 1520 GAAGACCATACAGCCAAAGCCATAAATCAACTGTGGAGTACACACTCT 1569
 434 heGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnGly 450
 1570 TTGAGGCGATGGAGTGGCCAGCGCTCCCTGGTGTGTCATCATGTCAGG 1619
 451 ArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGlyAr 467
 1620 AAGATTGTCTTGGAGTGGAAACATCAGTGTCTCAGCAAGGCGATGGCCG 1669
 467 gPheValProArgLysThrPheProAspPheValTyroLysArgIleLys 484
 1670 CTTCATCCCTCGAAGCCATTCAGGAGCATCTCTACCGCTGTCTCAGGA 1719
 484 laArgAsnArgLeuAlaGluIleHisGlyValProArgGlyLeuTyroAsp 500
 1720 TCAGAAGCAAGGTTTTCGGGTGTGATAGTGTTCAGGCGCATGTACAT 1769
 501 GlyProValHisGluValMetValProAlaLysProGlySerGlyAlaPr 517
 1770 GGGCTGTGTACGAGGTGGCAGCTACACCAACATGCTGCTCTCTCTCC 1819
 517 oAlaArgAlaSerCysProGlyLysIleSerValProProValArgAsnL 534

[illegible]

us-09-367-496-8.std.rng

Tue Jul 31 13:08:18 2001

```

548 p 54E
1788 T 1788

seq_name: /SIDS/gcgdata/geneseq/geneseq/NA2000.DAT: AAC40092
seq_documentation_block:
ID AAC40092 standard; DNA; 1734 BP.
XX
AC AAC40092;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 27016.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI0334C5-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

```

```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141847.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 1018.00
    Ratio: 2.844
    Percent Similarity: 68.451
    Length: 523
    Gaps: 9
    Percent Identity: 41.491

alignment_block:
US-09-367-496-8 x AAC40092
Align seg 1/1 to: AAC40092 from: 1 to: 1734

12 IleThrSerAspArgLeuIleArgGlyGlyArgIleValAsnAspAs 28

```

Tue Jul 31 13:08:18 2001

```

1018 .....CTAGGACATGGGAAGCCCTACAAGATGCCCTTTCCACAGCA 1059
325 AspLeuGlnValThrGlySerAlaHisCysThrPheThrAlaGlnLy 341
1060 ATTCCTCAGCTTGTAGAACTGATCACTGCTCACTTTCAATTCACACAAA 1109
341 sAlaValGlyLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyI 358
1110 AGCTCTAGGACTTGTATGATTTTCGCAAAATACCTAATGTTGTTAATGGCC 1159
358 leGluGluArgMetSerMetValTrpGluLysCysValAlaSerGlyLys 374
1160 TTGAGGAACCGATGCACTTCATATGGACACGATGCTGGAGTCTGGCCAA 1209
375 MetaspGluAsnGluPheValAlaValThrSerThrAsnAlaLysI 391
1210 CTCTCAGCTACTGATTAATGTTGCAATTAACCAACGATGCTGCTGCTAGAAT 1259
391 ePheAsnPheThrProArgLysGlyArgValAlaValGlySerAlaLysThr 424
1260 TTTCACATATATCCACGAAAGGAGCTATCTCTGCTGGCTGGATGCGAG 1309
408 sPLeuValIleTrpAsnProLysAlaThrLysIleIleSerAlaLysThr 424
1310 ATATTATCATATTGAATCCAACTCAAGCTACGAGATAGCTCAAGTCT 1359
425 HisAsnLeuAsnValGluTyArgAsnIlePheGluGlyValGluCysArgG 441
1360 CATCATCAAGATCAGACACAACTGCTACGAGGCGCAGAGGAAAGGG 1409
441 yAlaProAlaValIleSerGlnGlyArgValAlaLeuGluAspGlyL 458
1410 AAAAGTTCAAGTGACAATAGCAGGAGGAGCAATGTTGGGAAACACGAGG 1459
458 yMetPheValThrProGlyAlaGlyArgPheValProArgLysThrPhe 474
1460 AACTTAAAGTTGTTCCCAAGAGTGCAAGTATATA.....GAGATG 1500
475 ProAspPheValTyLysArgIleLysAlaArgAsnArgLeuAlaGluI 491
1501 CCTCTCTTCAGTTAC..... 1515
491 eHisGlyValProArgGlyLeuTyArgGly..... 501
1516 .....CTTTTCATGCTATTGAGAAATCAGATGCTA 1546
502 .....ProValHisGluValMetValProAla 510
1547 ATTATCTATCTCTCTTCGAGCTCCAGCTTAAGCGTGTGCAAACTGAAGCT 1596
511 LysProGlySerGly 515
1597 ACCTAAAGTGCAGGT 1611
seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:AAAT31258
seq_documentation_block:
ID AAT31258 standard; DNA; 1569 BP.
XX
AC AAT31258;
XX
DT 31-JAN-1997 (first entry)
XX
DE Hydatoninase coding sequence.
XX
KW Hydatoninase; hydantoin; d-N-carbamoyl-alpha-amino acid;
KW intermediate; drug synthesis; penicillin; cephalosporin;
KW antibiotic; ss.
XX
OS Pseudomonas sp.
XX
FH Key
FT CDS 60..1518

```

FT

PT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

/*tag= a
/product= Hydantoinase

W09620275-A1.

04-JUL-1996.

26-DEC-1995; 95WO-JP02688.

28-DEC-1994; 94JP-0326865.

(KANF) KANEGAFUCHI KAGAKU KOGYO KK.

Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
Yanada Y;WPI; 1996-321848/32.
P-ESDB; AAW03025.

Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
a recombinant hydantoinase derived from a strain of *Pseudomonas*,
Agrobacterium or *Bacillus*

Claim 19; Page 36-40; 54pp; English.

D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
hydantoin by treatment with a hydantoinase expressed by a
transformant microorganism carrying a vector containing DNA coding
for the hydantoinase and derived from *Bacillus* sp. KNR245.
Agrobacterium sp. KNR712 or *Pseudomonas* sp. KNR003A. The D-N-
carbamoyl-alpha-amino acid can be used for the production of
optically active alpha amino acids (especially D-phenylglycine and D-
p-hydroxyphenylglycine) as intermediates for drug synthesis,
especially for the production of semi-synthetic penicillin and
cephalosporin antibiotics.

Sequence 1569 BP; 327 A; 474 C; 485 G; 283 T; 0 other;

alignment_scores:

Quality: 1011.50 Length: 458
Ratio: 3.084 Gaps: 5
Percent Similarity: 71.616 Percent Identity: 45.852

alignment_block:

US-09-367-496-8 x AAT31258 ..

Align seq 1/1 to: AAT31258 from: 1 to: 1569

18 LeuileArgGlyGlyArgIleValAlaAspGlnSerPheTyrAlaAs 34
119 TATCCGTGATCGAAGCGGAAGAAATGCCCGATCGGAGGATCG... 164
51 alProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
165CAGGGCGACAAAGATTGTAGACGCGGAGGTGCCTATATCATTC 209
68 GlyGlyValAlaHisThrArgLeuGlnMetProValLeuGlyMetH 84
210 GCGCGCATGATCCTCACACGATCTTGAATGCCCTTTATGGGCACGAC 259
84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
260 CACTCGGACGACCTGGGATCGGACATTTCCGGCTCTCTCCGGCGGCA 309
101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeu 117
310 CCACGATGTTGTGATTTCTGTCATTCCTCCGCGGCG...GCGGGAATGCTT 356

118 AlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAs 134
357 GCGCGCTTCGATCAATGGCAGGAGGCGCGCGAGCTTCGTCGCA 406
134 pTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGlu 151
407 CTACTCGCTGCATATGTGCGTGACCGCGTGTCAAAAGCAGATCTTCGAG 456
151 LuLeuGluAlaLeuValLysGluLysGlyValAlaAsnSerPheLeuVal 167
457 ACATG...GCGAAGTGTGAGCGCGGTGTCAACACCTTCAAGCAATTC 503
168 MetAlaTyrIleLysAspArgCysGlnCysSerAspSerGlnMetTyrGlu 184
504 ATGCCCTATAAGGCGCGCTGATGGTGAACGACGACGAGATGTTCGCCTC 553
184 ePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
554 CTTCACGCGCTCGCGCGCGCTGGCGCGCTCCGCGCTCGTCATCGCGAAA 603
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
604 ACGGCGACATCGTCGCTCGTCACAGAAATATATGACGAGGAGCGCTC 653
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
654 ACGGGTCCGAGGCTCATGCTATTCGCCCGCGCGAGGTGCGAGGCGCA 703
234 uAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 251
704 GCGCACCAACCGCGCATCATGCTCCGATCCGACGCGCTCGCGTCT 753
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
754 ATATCGTGATCTCTCTGTGAACAGGCGCAGCAAGCATCGCGCGCGCA 803
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
804 CGCCAGAAGGAATCGCGCTATGGCGGCGCTCATCCAGCACCTGCT 853
284 yThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaLap 301
854 GCTTGACGAGTCCGAGTACAAGAGAGCGGCTGGGACGAGCGCGCTCGGC 903
301 heValThrSerProProValAsnProAspProThrThrAlaAspHisLeu 317
904 GGGTATGTCGCGCCATTCGC...GATAAAGCCATCAGGACAGCGCTC 950
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
951 TGGGCTGGGCTCGCAGCGGCTCGCTCAGGTCTGGCAACGGNCCACTG 1000
334 sThrPheThrThrAlaGlnLysAlaValGlyLysAspAsnPheAlaLeu 351
1001 CGCCTTTACCACTGAACGAGGCGGCTGGGACTCAACGACTTCACGAAA 1050
351 leProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrpGlu 367
1051 TCCGAAACGCGACCGGGGGTTAGAGGACCGGTGTGCTGCTGCTGAGAC 1100
368 LysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaValTh 384
1101 TATGGGTGAAGACGGGGGCTCAGCGCTTAACGAATTCGTCGCGCTCAC 1150
384 rSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyArg 401
1151 CTCACCAATATCGCAAAATCTGAACATCTATCCGCAAAAGGCGCAA 1200
401 aAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAlaThr 417
1201 TCTGCTGCTGGTCCGATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1250
418 LysIleSerAlaLysThrHisAsnLeuValGluTyrAsnIlePh 434

34 pValHisValGluAspGlyLeuIleLysGlnIleGlyClnAsnMetLeuIleV 51
::: :::::::::::::: ||| |
404 TCTTCATCAATTCAAGATGGGAATAATGCAGTAAATCGGGAATAATTA...G 450
:
51 aProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
::: :::::::::::::: ||| |
451 ATCAGAGCGGACGGGAAGTGATTGATGCCACAGGTGTGTTATGCTGTTCCA 500
:
68 Gly3YvalAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
||| ||| :::: ||| |
501 GGA3GCATTTGATCCGCACACCATTAGATATGCGGTTGCGGCACACTGT 550
:
84 rProAlaAspAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
::: :::::::::::::: ||| |
551 GACAAAAGACGACTTGCTGCGGGACGATTGCCGCCCATTTGGCGSGA 600
:
101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
||| ||| :::: ||| |
601 CGAGCACCATATTGATTTTTCTTAAACGATAAAGGTGAGCCCTGAAA 650
:
118 AlaAlaTyrcLuGlInrPrArgGluArqAlaAspSerAlaAlaCysCysAs 134
||| ||| :::: ||| |
651 AAAGCGATTGAACTTGGCATTAACAAGCGACGGGAANGCGGTATCGA 700
:
134 pTyrSerLeuHisValAspIleThrargTrpHisGluSerIleLysGluG 151
||| ||| :::: ||| |
701 TTAGCGGTTCCATTGATGATCAGTGAATAACGGACGATGCTGTGAAG 750
:
151 luLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPhe 167
||| ||| :::: ||| |
751 AGCTTCCAAAAGTGATCGAGAAGAAGAAATTACCTCTTTAAAGATT 800
:
168 MetAlaTyrlLysAspArgCysGlnCysSerAspSerGlnMetYrGluIl 184
||| ||| :::: ||| |
801 ATGCGGTATAAGATGTGTTCAAGCTGATGATGGAACCTTGTATCGGAC 850
:
184 ePheSerIlelEargAspLeuGlyAlaLeuAlaGlnValHisAlaGluA 201
::: :::::::::::::: ||| |
851 GCTAGTCGCGCAAAAGAACCTCGAGCGCTGTGTCATGTGTCATGCCGAGA 900
:
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluLeuGlyIle 217
||| ||| :::: ||| |
901 ATGGAGACGTGATTGACTATTTAACGNAAAAAGCCTTGGAGGACGGCAT 950
:
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaGl 234
||| ||| :::: ||| |
951 ACTGATCCGATTTATCATGCTAATTAACGAGACCTCCAGAGTAGTAGGAGA 1000
:
234 uAlaValTyrgAlaValThrIleAlaLysGlnAlaAsnCysProLeut 251
||| ||| :::: ||| |
1001 AGCGACGGGCGGCCCTGTCAATTGACAGAACTCGCTGGTTCCCAATGCT 1050
:
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
||| ||| :::: ||| |
1051 ACSTCGTTCACTGATCTCGTGTCTCAACGCGGTAGAGAAAATTGCTGAAGCG 1100
:
268 LysArgArgGlyValValPheGlyCluProIleThrAlaSerLeuGl 284
||| ||| :::: ||| |
1101 CGCAATAAGGGGTTGAATGATGCGGGCGAACTGTGCCAGTATCTGGT 1150
:
284 yThrAspGlySerHisTyrrTpSerLysAsnTrpAlaLysAlaAlaAlap 301
||| ||| :::: ||| |
1151 GCTCGATCAGTCCCTATTAGAAAACCGCAATTT...GAAGGTGCTAAAT 1197
:
301 heValThrSerProProValAsnPProaspProThrThrAlaAspHisLeu 317
||| ||| :::: ||| |
1198 ATGSTATGGTCACCGCGCTTCGT...GAGAAATGGCATCAAGAAGTGCCTA 1244
:
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisCy 334
||| ||| :::: ||| |
1245 TGSAATGTGCTTGAAAACGGCCAGCTGCAACGCTCGGATCTGACCAANTG 1294
:
334 sThrPheThr...ThrAlaGlnLysAlaValAlcLyLysAspAsnPheAlaL 350

[illegible]

```

1295 CTCAATTGATTTAAAGCCAAAAGATATTAGGAAGGGAGATTTTACCA 1344
350 euileProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
1345 AAATCCCAATGGGTCCTATTATTAGGATCGGGTGGATTTCTTTTC 1394
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVa 383
1395 AGTGAAGGAGTGAAAGGAGAGATACCTCAACAGTTTGTGTATAT 1444
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyProArgLysGlyA 400
1445 TGTATCAACAAGATCCCAAAATTTGTTGGTCTATTCCGAAGAAAGGAA 1494
400 rgValAlaValGlySerAspAlaAspLeuValIleTyrAsnProLysAla 416
1495 CCATTCGGTCGGTCGGATCGGATTTAGTCAATTTTGTATCAACGGTT 1544
417 ThrLysIleleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnI 433
1545 GAACGGGTGATTTTCAGCCGAACACACCATATGGCTGTGGATTATATCC 1594
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerClnG 450
1595 GTTGAAGGATGAAGATACAGGGGAACCTGTGCGTTTATGTAGAG 1644
450 lyArgValAlaLeuGluAspGlyLysMetPheValThrProGlyValAlaGly 466
1645 GAGAAATTTGTTGACGTGATAAAATTTTCGCGAAACCGGGGTACGGC 1694
467 ArgPheValProArgLysThrPhe 474
1695 CAATATGTTAAACCGCGCAATAT 1718

```

seq_name: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT:AAQ80565

seq_documentation_block:

```

ID AAQ80565 standard; DNA; 1383 BP.
XX
AC AAQ80565;
XX
DT 27-OCT-1995 (first entry)
XX
DE B.thermoglucoasius D-hydantoinase gene with engineered mutation.
XX
KW D-hydantoinase; dihydropyrimidinase; muten; variant; mutation;
KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;
KW production; racemisation; ss.
XX
OS Bacillus thermoglucoasius.
XX
FH Key Location/Qualifiers
FT CDS 1..1383
FT /tag= a
FT /EC_number= 3.5.2.2
FT /note= "Deletion of a single base in the wild-type
FT ORF results in this sequence coding for an
FT enzyme which is 12 amino acids shorter than
FT the wild-type protein and the last 6 amino
FT acids are different"
XX
PN DB4328829-A.
XX
XX
XX 02-MAR-1995.
XX
XX 27-AUG-1993; 93DE-4328829.
XX
XX 27-AUG-1993; 93DE-4328829.
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
XX Burtcher H, Lang G, Popp F;

```

```

XX WPI; 1995-099331/14.
DR P-PSDB; AAR69874.
XX
XX New highly active and stable D-hydantoinase enzyme - isolated
PT from Bacillus thermoglucoasius, useful in the prodn. of
PT D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin
XX
XX Claim 3; Page 7; 8pp; German.
XX

```

The wild-type D-hydantoinase gene was isolated from *Bacillus thermoglucoasius* using the amplification primers Hyd1 and Hyd2 (see AAQ80566 and AAQ80567). A single-base deletion in the HindIII site (from AAGCTT to AAGCT) was introduced to give a coding region (AAQ80565) that encoded a D-hydantoinase (AAR69874) that was shorter than the wild-type protein with a different C-terminal sequence (see Features Table). The mutant enzyme was found to have improved heat stability and higher activity compared to known hydantoinases.

Sequence 1383 BP; 427 A; 251 C; 368 G; 337 T; 0 other;

alignment_scores:

Quality: 826.00 Length: 458
Ratio: 2.614 Gaps: 4
Percent Similarity: 68.996 Percent Identity: 39.738

alignment_block:

US-09-367-496-8 x AAQ80565

Align seg 1/1 to: AAQ80565 from: 1 to: 1383

```

18 LeuIleArgGlyGlyArgIleValAsnAspGlnSerPheTyrAlaAs 34
   :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 ATATAAAATGGAACGATTTACCGCAACCGCATACGTATGAAGCGGA 59
   :::::|||||:|||||:|||||:|||||:|||||:|||||:
34 pValHisValGluAspGlyLeuIleLysGlnIleGlyGluAsnLeuIleV 51
   :::::|||||:|||||:|||||:|||||:|||||:|||||:
60 CTGTGCTCATTAAGACGGAAATTTGCCATGATAGGCCAACATTTA...G 106
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
51 alProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
107 AAGAAAAGCGCGTGAAGTGATGATGCCAAAGCGCTACGTATTTCCTCA 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMetTh 84
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GCGGTATTGATTCGCACACGATTTAGATATGCCCTTTGGCGGCACGGT 206
   :::::|||||:|||||:|||||:|||||:|||||:|||||:
84 rProAlaAspAspPheCysGlnGlyThrLysAlaLaLeuAlaGlyGlyT 101
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 GACAAAGGATGATTTTCAATCTGGAACGATTGCGGCGGCATTGGCGGAA 256
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 CAACGACCATCATCGACTTTTGTAAACGAATAAAGGGGAGCCATTAAAA 306
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
118 AlaAlaTyrGluGlnTrpArgGluArgAlaAspSerAlaAlaCysCysAs 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
307 AAAGCGATTGAACTTGGCACACAAAGCGAAGGAAAGCGGTATTTCGA 356
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
134 pTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleLysGluG 151
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 TTATGGCTTCCATTTAATGATTAGCGAAATACGGATGACGTATTAGAAG 406
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 luLeuGluAlaLeuValLysGluLysGlyValAsnSerPheLeuValPhe 167
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 AGCTGCCAAAAGTCATTCCCGAAGAGGAGTAACATCTTTAAAGTGTTT 456
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 MetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
457 ATGGCGTATAAAACGATTTTCAGGCAGATGATGGAACGTTATATACCGC 506
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 ePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGluA 201

```

Tue Jul 31 13:08:18 2001

507 GCTAGTGGCTGCCAAGAACTTGGCGCGCTTGTCATGTTTCATGCGGAAA 556
::: :::
201 snglyaspIleValGluGluGluGlnLysArgLeuLeuGluLeuGlyLe 217
::: :::
557 ATGGGGATGTGATTGATTACTTACGAAAAAAGCGCTTGGCGAAGGAAT 606
::: :::
218 ThrGlyProGluGlyHisValLeuSerHisProGluGluValGluAlaG 234
::: :::
607 ACGGAGCCGATTTACCATGCTTTTAAACCGCGCTCCAGAAATAGAAAGGA 656
::: :::
234 uAlaValTyrArgAlaValThrIleAlaLysSclAlaAsnCysProLeu 251
::: :::
657 AGCGACCGGCGCGCTGTCAATTGACAGAGCTTGCCGGTTCACAACCTTT 706
::: :::
251 YrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAla 267
::: :::
707 ACGTGTTCACGTGACATGTCGCAAGCGGTGGAAAAAATTTGCCAAGCG 756
::: :::
268 LysArgArgGlyValValValPheGlyGluProIleThrAlaSerLeu 284
::: :::
757 CGCAATAAAGGTTGGATGTCGGGGAAGAAAGCTCCGCAATATCTTGT 806
::: :::
284 YThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaIap 301
::: :::
807 TCTCGACCAATCG...TATTAGAAAAGCCTCATTTTGAAGGCGCGAAAT 853
::: :::
301 heValThrSerProProValAsnProAspProThrThrAlaAspHisLeu 317
::: :::
854 ATGTTGGTGCCTCCGCTCGCT...GAAAAATGGCATCAAGAAGTATGTG 900
::: :::
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
::: :::
901 TGGATCGCGTGAAAAACGCCAGCTGCAACGCTTGGATCGACCAATG 950
::: :::
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAlaL 350
::: :::
951 TTCAATTGACATTTAAAGGCCCAAAAGAACTTGGCAGAGGAGATTTACTA 1000
::: :::
350 euIleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
::: :::
1001 AATYTCCAACGGGGCGCCGATGGTCGAGGATCGGGTCAGCATCTCTTTC 1050
::: :::
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVa 383
::: :::
1051 AGTGAAGGGTTAAAAAAGGAAGAATCACGTTAAATCAATTTGTCGATAT 1100
::: :::
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyA 400
::: :::
1101 TATGTGCAAGAATTTGCCAAATTTGTCGGGTATTCCCGAGAAAAAGGA 1150
::: :::
400 rGValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
::: :::
1151 CGATCGCGGTAGGTTTCAGACGACGACTAGTCAATTTTGTACCCCGGATATC 1200
::: :::
417 ThrLysIleIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnI 433
::: :::
1201 GAACGGGTGATTTCCGGGAACAACCATATGCGCTCGACATATATATGC 1250
::: :::
433 ePheGluGlyValGluCysArgGlyAlaProAlaValIleSerGlnG 450
::: :::
1251 ATTTGAAGGAATGAAAGTAACGGGTGAACCGGTATCGGTTCTGTGCAGAG 1300
::: :::
450 lYArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
::: :::
1301 GCGAATTTGTTCTCGGTGATAAACAATTTGCGGAAACACAGGGTACGGC 1350
::: :::
467 ArgPheValProArgLysThrPhe 474
::: :::
1351 CAATATTTAAAGCTTGGCTGTTTT 1374
::: :::

seq name: /SIPDS8/qcdata/qeneseq/qeneseq/NA1987.DAT: AAN70154

```

seq_documentation_block:
ID      AAN70154 standard; DNA; 1883 BP.
XX
XX      AAN70154;
XX
XX      AC      (first entry)
XX      DT      1220.
XX      DE      Hydatonase gene contg. construct Lu 1220.
XX      KW      Construct Lu 1220; mesophilic microorganism; thermophilic hydantoinase;
XX      KW      ss.
XX      OS      CBS 303.80.
XX      XX      EP219034-A.
XX      PN      22-APR-1987.
XX      PD      09-OCT-1985; 85DE-3535987.
XX      PR      (BADI ) BASF AG.
XX      PA      Jacob E, Henco K, Marcinowski S, Schenk G;
XX      PI      WPI; 1987-110013/16.
XX      DR      prodn. of mesophilic microorganisms contg. thermophilic hydantoinase -
XX      PT      by genetic modification with DNA from thermophilic microorganisms.
XX      PT      Disclosure; Fig. 12A-B; 27pp; English.
XX      PS      The construct Lu 1220 contains a DNA sequence from a thermophilic
XX      CC      microorganism which cleaves D-hydantoin. The DNA fragment encoding
XX      CC      hydantoinase comprises nucleotides 391-1746.
XX      CC      The enzyme is useful for seroselective conversion of D,L-hydantoin to D-alpha
XX      CC      D,N-carbamoyl-alpha amino acids, which are intermediates for D-alpha
XX      CC      amino acids. These are useful in the prodn. of penicillins and
XX      CC      cephalosporins. Transformed mesophilic microorganisms, esp. E.coli.
XX      CC      congly. the DNA can produce 4-40 times greater enzyme activity than
XX      CC      wild type organisms. The DNA is from a thermophilic organism, esp.
XX      CC      CBS 303.80 or 363.80. The cloning vector is linearised cosmid or
XX      CC      lambda phage DNA.
XX      CC      See also AAN70153.
XX
XX      Sequence 1883 BP; 592 A; 340 C; 482 G; 469 T; 0 other;
XX

```

alignment_scores:		
Quality:	805.00	Length: 458
Ratio:	2.547	Gaps: 4
Percent Similarity:	68.996	Percent Identity: 39.738

alignment_block:

alignment_block:
IIS-09-367-496-8 X AAN70154

Align seg 1/1 to: AAN70154 from: 1 to: 1883

18 LeuIleArgGlyGlyArgIleValAsnAspAlcInserPheTyrAlaAs 34
:::|||||:::|||||:::|||||:::|||||
400 ATATATAAAATGGACCATGTTCACCGCAACGATATCATGAAGCGA 449

34 pValHisValGluAspGlyLeuIleLysGlnIleGlyGluasnLeuIleV 51
||:::|||||:::|||||:::|||||:::|||||
450 CTTGTCTATTAAAGACGGAAAATTCGCATGATAGGCCAACATTTA...G 496

51 alProGlyGlyIleHisThrIleAspAlaHisGlyLeuMetValLeuPro 67
:::|||||:::|||||:::|||||:::|||||
497 AAAAAAAGGGCGCTGAAGTGATTGATGCCAAAGCGCTTCATCGTATTCCA 546

68 GlyGlyValAspValHisThrArgLeuGlnMetProValLeuGlyMethTh 84
CCCCCCTTCCGACACCGCATTTTACATATGCGTTGTGGCGCACGGT 996


```

84 rProAlaAspPheCysGlnGlyThrLysAlaAlaLeuAlaGlyGlyT 101
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
597 GACAAAGGATGATTCGAATCGGAACGATTCGGCGGCGATTCGGCGAA 646
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 hrThrMetIleLeuAspHisValPheProAspThrGlyValSerLeuLeu 117
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
647 CACCGACCATCATCGACTTTTCTTTAACGAATAAAGGGGAGCCATTAAAA 696
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 AlaAlaTyrGluGlnTrpArgGluAlaAspSerAlaAlaCysCysAs 134
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
697 AAAGCGATTGAACCTGGCACACAAAGCGAAGGAAAGCGTTATTGA 746
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 pTyrSerLeuHisValAspIleThrArgTrpHisGluSerIleGluG 151
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
747 TTATGGCTTCCATTTAATGATTAGCGAAATACGATGACGTATTAGAAG 796
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 LuLeuGluAlaLeuValLysGluLysGlyValAlaSerPheLeuValPhe 167
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
797 AGCTCCCAAGATCATTCGCGAAGGAGGATAACATCTTTAAAGTGTTT 846
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
168 MetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMetTyrGluI 184
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
847 ATGGCGTATAAAACGATTTTCAGGCAGATGATGAACGTTATACCGCAC 896
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 ePheSerIleIleArgAspLeuGlyAlaLeuAlaGlnValHisAlaGlu 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
897 CTAAGTGGCTCCAAAGAACCTGGCGGCTTGTCTATGTTTCATCGGAAA 946
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 snGlyAspIleValGluGluGlnLysArgLeuLeuGluGlyIle 217
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
947 ATGGGATGCTGATTACTTACGAAAGGCGCTTCGGAAGGGAAT 996
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 ThrGlyProGluGlyHisValLeuSerHisProGluValGluAlaG 234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
997 ACGAGCGCGATTTACCATGCTTTAACGCGGCTCCAGAAGTAGAAGGAGA 1046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 uAlaValTyrArgAlaValThrIleAlaLysGlnAlaAsnCysProLeu 251
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1047 AGCGACGGCGCGCTGTCAATGACGAGCTTCGCGGTTCACAACTTT 1096
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 yrValThrLysValMetSerLysGlyAlaAlaAspAlaIleAlaGlnAl 267
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1097 ACGTTGTTCACGTGACATGTGCGCAAGCGGTGGAAAAAATTCACACAGC 1146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 LysArgArgGlyValValPheGlyGluProIleThrAlaSerLeuG 284
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1147 CGCAATAAAGGTTGGATGTGGGAGAACCGTTCGCGCAATATCTTGT 1196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 yThrAspGlySerHisTyrTrpSerLysAsnTrpAlaLysAlaAlaAl 301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 TCTCGACCAATCG...TATTTAGAAAAGCTGTATTTGAAGCGCGCAAT 1243
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 heValThrSerProValAsnProAspProThrThrAlaAspHisLeu 317
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1244 ATGTTTGGTCCCTCCGCTCGF...GAAAAATFGGATCAAGAGTATTG 1290
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
318 ThrCysLeuLeuSerSerGlyAspLeuGlnValThrGlySerAlaHisC 334
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1291 TGGAAATCGCTGAAAAACGCGCAGCTGCACACGCTTGGATCGGACCAATG 1340
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334 sThrPheThr...ThrAlaGlnLysAlaValGlyLysAspAsnPheAla 350
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1341 TTCATTGTGACTTTAAAGGCCAAAAAGAACTTGGCAGGAGAGATTTACTA 1390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 euIleProGluGlyThrAsnGlyIleGluGluArgMetSerMetValTrp 366
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1391 AAATTCCAAACGGCGCGCGATGTCGAGGATCGGGTCAGCATTTCTTTC 1440
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 GluLysCysValAlaSerGlyLysMetAspGluAsnGluPheValAlaVa 383
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1441 ACTGAAGGGGTTAAAAAAGGAAGATACGTTAAATCAATTTGTTCATAT 1490
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 lThrSerThrAsnAlaAlaLysIlePheAsnPheTyrProArgLysGlyA 400

```

```

: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1491 TATGTCGACAAGAAATTCGCAAAATGTTGGGTATTCCCGAGAAAGGAA 1540
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 rgValAlaValGlySerAspAlaAspLeuValIleTrpAsnProLysAla 416
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1541 CGATCGCGTAGGTTTCAGACGAGACTTAGTCATTTTACCGCGGATATC 1590
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 ThrLysIleSerAlaLysThrHisAsnLeuAsnValGluTyrAsnIle 433
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1591 GAACGGGTGATTTCGCGGAAACACACCATATGCGCGTCGACTATATGC 1640
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 ePheGluGlyValGluCysArgGlyAlaProAlaValValIleSerGln 450
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1641 ATTGAAGGAATGAAAGTAAACGGGTCAACCGGTATCGGTTCG.TGCAGAG 1689
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 lYArgValAlaLeuGluAspGlyLysMetPheValThrProGlyAlaGly 466
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1690 GCGAATTTGTTGTCGTTGATAACAATTTGTCGGAACCCAGGTTACGGC 1739
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 ArgPheValProArgLysThrPhe 474
||||| |||||: |||: |||: |||: |||: |||: |||: |||: |||:
1740 CAATATTTAAACCGGCAAAATAC 1763
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/NAI999.DAT:AAI13316

seq_documentation_block:

ID AAI13316 standard; DNA; 13085 BP.

XX AC AAI13316;

XX DT 19-MAR-1999 (first entry)

XX DE Enterococcus faecalis genome contig SEQ ID NO:379.

XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;

XX KW vaccine; attenuation; computer readable medium; ds.

XX OS Enterococcus faecalis.

XX PN WO9850555-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08985.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Dillon PJ, Kunsch CA;

XX DR WPI; 1999-045171/04.

XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides

XX PT - used to develop products for the detection of Enterococcus and for

XX PT use in vaccines for prevention or attenuation of Enterococcus

XX PS infection.

XX PS Claim 1; Page 1549-1556; 2084pp; English.

XX CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAI12938 to AAI13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as contigs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the

Tue Jul 31 13:08:18 2001

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 13085 BP; 3789 A; 2961 C; 2106 G; 4222 T; 7 other;

alignment_scores:
Quality: 698.50 Length: 462
Ratio: 2.268 Gaps: 9
Percent Similarity: 66.667 Percent Identity: 34.199

alignment_block:

US-09-367-496-8 x AAX13316/rev ..

Align seg 1/1 to reverse of: AAX13316 from: 1 to: 13085

```

17 LeuLeuileAArgGlyGlyArgileValAsnAspGlnSerPheTyrAl 33
   ::::::::::::::::::::
8239 ATTCTGCTCAAGGTGGAACCGTCTGATCTGCATATAATCGTCAATT 8190
   ::::::::::::::::::::
33 aaspValHisValGluAaspGlyLeuileGlyGlnleGlyGluAsnLeuI 50
   ::::::::::::::::::::
8189 GGAATTCGAATTGACGGCGAGCAAAATTGTAGAGATGGGGCAAAATTAC 8140
   ::::::::::::::::::::
5C leValProGlyLeuHisThrileAspAlaHisGlyLeuMetValLeu 66
   ::::::::::::::::::::
8139 CAGTAGAAATCT...CAGATTGAGGATGTTACGGGTGTTATCTATTA 8093
   ::::::::::::::::::::
67 ProGlyGlyValAspValHisThrArgLeuGlnMet.....ProValle 81
   ::::::::::::::::::::
8092 CCAGGCTTTATGATCGCATACGATCTGGAATTAATAATGCAAGG 8043
   ::::::::::::::::::::
81 uGlyMetThrProAlaAspPheCysGlnGlyThrLysAlaAlaLeuA 98
   ::::::::::::::::::::
8042 TTCGTTAAGTACGGCGGATTAACCTACCTGGAACCAAGCGCGTGG 7993
   ::::::::::::::::::::
98 laGlyGlyThrThrMetileLeuAspHisValPheProaspThrGlyVal 114
   ::::::::::::::::::::
7992 CCAAGGTACGACTACGGTTATTGATGATGCTACACCTAATAAGCGCG 7943
   ::::::::::::::::::::
115 SerLeuLeuAlaAlaTyrGluGlnTTPArgGluArgAlaAspSerAlaI 131
   ::::::::::::::::::::
7942 TCATTGAAAGACGTGTTAGTACTCTGGAATCAATGCTGCAAGCAAG 7893
   ::::::::::::::::::::
131 aCysCysAspTyrSerLeuHisValAspIleThrArgTTPHisGluSerI 148
   ::::::::::::::::::::
7892 TTCATGTGATTTACTTACCATATGTCATGATGATGATGAAACCAACGA 7843
   ::::::::::::::::::::
148 leLysGluGluLeuAlaLeuValLysGlyLysGlyValAsnSerPhe 164
   ::::::::::::::::::::
7842 TAGTGTGTAATTCAGGAATGATT...GCTGCAGGTATATCTCTTC 7796
   ::::::::::::::::::::
165 LeuValPheMetAlaTyrLysAspArgCysGlnCysSerAspSerGlnMe 181
   ::::::::::::::::::::
7795 AAATGTATATGGCTTAT...GATAATTGGGACCAACAGATGCGGAAT 7749
   ::::::::::::::::::::
181 tTyrGluLlePheSerIleleAArgAspLeuGlyAlaLeuAlaGlnValH 198
   ::::::::::::::::::::
7748 TTTTGAAGCGATGAAGGAATAAATAAGTTAATGGAATGTTAGCGGTT 7699
   ::::::::::::::::::::
198 IsAlaGluAsnGlyAspIleValGluGluGlnLysArgLeuLeuGlu 214
   ::::::::::::::::::::
7698 APTGTGAATGTTGGTATTAGTAGATGAATTAATCCATCATATGTTCT 7649
   ::::::::::::::::::::
215 LeuGlyIleThrGlyProGlyHisValLeuSerHisProGluGluVa 231
   ::::::::::::::::::::
7648 CAAAGAAAGCTAACCCGACATCATCGTTATCAAGACCGCGCGCTGT 7599
   ::::::::::::::::::::
231 lGluAlaGluAlaValTyrArgAlaValThrileAlaLysGlnAlaAsnC 248
   ::::::::::::::::::::
7598 TGAAGCAGAAAGCGGTAGCAAGATATTGATGATGATGCGGAATGCTGAT 7549

```

```

248 ysProLeuTyrValThrLysValMetSerLysGlyAlaAlaAspAlaile 264
   ::::::::::::::::::::
7548 TGCCGCTTAAATATTGTCATTTAAGCACAACAGATCATTAAGAGCGGTG 7499
   ::::::::::::::::::::
265 AlaGlnAlaLysArgArgGlyValValPheGlyGluProIleThrAl 281
   ::::::::::::::::::::
7498 GAACGGGCTCGTCAAGAGAGCAATCAGTCTATGTAGAAACTGCGCTCA 7449
   ::::::::::::::::::::
281 aserLeuGlyThrAspGlySerHisTyrTrpSerLysAsnTTPAlaLysA 298
   ::::::::::::::::::::
7448 ATATCTTTTACTAGATGATCATTGTATGATCGCTCAATTTT...GAAG 7402
   ::::::::::::::::::::
298 laAlaAlaPheValThrSerProValAsnProAspProThrThrAla 314
   ::::::::::::::::::::
7401 CAGCAAAATATGCTGCTCACCGCCTACTACG...AAACGA 7364
   ::::::::::::::::::::
315 asHisLeuThrCysLeu.....LeuSerSerGlyAspLeuGlnVa 328
   ::::::::::::::::::::
7363 GAGGATCAACGGGCTTTGTGCAAGGATTAAGAGGGGCAATCAATAC 7314
   ::::::::::::::::::::
328 lThrGlySerAlaHisCysThrPheThr...ThrAlaGlnLysAlaValG 344
   ::::::::::::::::::::
7313 AATATCAGCAGATCAGTCTGATTTTAAATTTTACGGACAAAAACAGTGTG 7264
   ::::::::::::::::::::
344 lLysAspAsnPheAlaLeuIleProGluGlyThrAsnGlyIleGluGlu 360
   ::::::::::::::::::::
7263 CCAAGATGATTTTGTAGTAAATTTCCCAATGGATGCCAGGTGTTCAACG 7214
   ::::::::::::::::::::
361 ArgMetSerMetValTTPGluLysCysValAlaSerGlyLysMetAspG1 377
   ::::::::::::::::::::
7213 AGACGAGATTAATTTACACGGAAGGTGTGCGAAGGCGGCAATCACCTT 7164
   ::::::::::::::::::::
377 uasnGluPheValAlaValThrSerThrAsnAlaAlaLysIlePheAsnP 394
   ::::::::::::::::::::
7163 GGAATAAATGTTGCTGCTTCTGAAATATTTGCTAAACAATTTCTGA 7114
   ::::::::::::::::::::
394 heTyrProArgLysGlyArgValAlaValGlySerAspAlaAspLeuVal 410
   ::::::::::::::::::::
7113 TGTATCCCAAAAAGGTGTGTCGAAGAGGCGAGTGTGCGCATTTAGTT 7064
   ::::::::::::::::::::
411 IleTTPAsnProLysAlaThrLysIleleSerAlaLysThrHisAsnLe 427
   ::::::::::::::::::::
7063 GTTGGGATCCCAACCAACCGCGGTGATGCTGCGAGAAGCAGTTACA 7014
   ::::::::::::::::::::
427 uasnValGluTyrAsnIlePheGluGlyValGluCysArgGlyAlaProA 444
   ::::::::::::::::::::
7013 AATATGCGATTATACCGCTATGAGGATTTGAAACACCAAGGACAGGCTC 5964
   ::::::::::::::::::::
444 laValIleSerGlnGlyArgValAlaLeuGluAspGlyLysMetPhe 460
   ::::::::::::::::::::
6963 GAATGTTTATTAAAGAGGTCAAAAAGTTGCTCAAGCTGGCAAGTC... 6917
   ::::::::::::::::::::
461 ValThrProGlyAlaGlyArgPheValProArgLys 472
   ::::::::::::::::::::
6916 ATCTGCAAAATCAAGGAAATTTGTTTTCGTAAA 6881
   ::::::::::::::::::::
seq_name: /SID8/gcgdata/geneseq/geneseq/NA1995.DAT:AA04198
seq_documentation_block:
ID AA04198 standard; DNA; 1374 BP.
XX
AC AA04198;
XX
DT 13-MAR-1996 (first entry)
XX
DE Agrobacterium radiobacter D-hydantoinase gene.
XX
KW D-hydantoinase; D-N-carbamylase; enzyme; stereospecific reaction;
KW D-amino acid; ss.
XX
OS Agrobacterium radiobacter.
XX
FH Location/Qualifiers

```


CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 3039 BP; 744 A; 711 C; 659 G; 919 T; 6 other;

alignment_scores:

Quality: 600.00 Length: 166
Ratio: 4.225 Gaps: 0
Percent Similarity: 85.542 Percent Identity: 68.072

alignment_block:

US-09-367-496-8 x AAF21851 ..

Align seg 1/1 to: AAF21851 from: 1 to: 3039

406 AspAlaAspLeuValIleTrpAsnProLysAlaThrLysIleIleSerAl 422
|||||
3 GATGCCGACCTGGTCTATCTGGGACCCGACACGCTTAAAGCATCTCTGC 52
422 aLysThrHisAsnLeuAsnValIleGlyTrpAsnIlePheGluGlyValGluC 439
|||||
53 CAAGACACACAACAGCTCTCTCGAGTACACATCTTTGAGGCGATGGAGT 102
439 ySArgGlyAlaProAlaValIleSerGlnGlyArgValAlaLeuGlu 455
|||||
103 GCCGGGCTCCCCACTGGTGTGTATCATGCCACGAGGGGAAGATTGCTCGAG 152
456 AspGlyLysMetPheValThrProGlyAlaGlyArgPheValProArgLy 472
|||||
153 GACGGCACCTGCATGTNACCGAANGCTCTGGACGCTACATTCCTCCCGAA 202
472 sThrPheProAspPheValTyrLysArgIleLysAlaArgAsnArgLeuA 489
|||||
203 GCCCTTCCTGATTTTATTACAGCGTATCAGGCAAGGAGCAGGCTGG 252
489 laGluIleHisGlyValProArgGlyLeuTyrAspGlyProValHisGlu 505
|||||
253 CTGAGCTGAGAGGGTTCCTCGTGGCTGTATCAGGACCCCGTGTGTGA 302
506 ValMetValProAlaLysProGlySerGlyAlaProAlaArgAlaSerCy 522
|||||
303 GTGCTGTGAGGCCCAAGACAGTCACTCCAGCCTCCTCGGCCAAGACGTC 352
522 sProGlyLysIleSerValProProValArgAsnLeuHisGlnSerGlyP 539
|||||
353 TCCTGCCAAGCAGCGCCCGCCCTCTCCGGAACCTGCACCATCTGTGAT 402
539 heSerLeuSerGlySerGlnAlaAspHisIleAlaArgThrAla 555
|||||
403 TCAGTTGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATG 452
556 GlnLysIleMetAlaProProGlyGlyArgSerAsnIleThrSerLeu 571
|||||
453 CAGCGTATCGTGGCGCCCGCCCGTGGCGGTGCGCAACATCACACGCTG 500

Db 4 IIKNGTIVTADTYEADLLIKDKIAMIGQHL-EEKGAEVIDAKGCVYFPGGIDSHTHLD 62
QY 78 MPVLGMPADDFCQCTKAALAGTTMILDHVPDPTGVSLAAYEQRERADSAACCDYSL 137
Db 63 MPFGTGTVDKDDFESCTIAAAGGTTIIDFCLTNKGEPLKKAETWHNKAKGKAVIDYGF 122
QY 138 HVDITRWHESKEELEALVKEGVNSFLVFMAYKDRCCSDSQMSYEIFSIIRDLGALAQV 197
Db 123 HLMISEITDDVLEELPKVIAEGITSEKVFMAKYNVFOADDGTLRTLVAAKELGALVMV 182
QY 198 HAENGDIIVEEQKRLLELIGITGPEGHVLSHPEEVEAEAVYRAVTIAKOANGLYVTKYMS 257
Db 183 HAENGDIVIDLTKKALAEAGNTEPIYHALTRPEVEGEATGRACQTLTAGSOLYVHVTC 242
QY 258 KGAADATAQAKRRGVVFGPEITASLGTDGSHYWSKNWAKAAAFVTPSPVNPDPPTADHL 317
Db 243 QAAVEKTAQARNKGLDVMGETCPOYLVDQS-YLEKPDFEGAKYVWSPPLR-EKWHQEV 300
QY 318 TLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPGTTNGIERMSVMWEKCVASGKMD 376
Db 301 WNALKNGQLTGLSDQCSDFKQKELGRGDTKIPNGGPMVEDRVSLFSEGVKKGRIT 360
QY 377 ENFVAVTSTNAAKIFNFYPRKGRVAVGSDADLVWNPATKIISAKTHNLNVEYNIFEG 436
Db 361 LNQFVDMSTRIAKLFLGFPKGTIAVGSADLVIFDPDIERSVISAETHHMAVDYNAFEG 420
QY 437 VECRGAUVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 421 MKVTGEPVSLCRGEFVVRDKQFVGKPGGYLKAGCF 458

RESULT 2

US-08-602-656-1
; Sequence 1, Application US/08602656
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process For the
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
US-08-602-656-1

Query Match

Best Local Similarity 27.7%; Score 826; DB 1; Length 460;

Matches 182; Conservative 78; Mismatches 194; Indels 4; Gaps 4;

QY 18 LIRGRIYNDQDSFYADVHVEDGLIKIGENLIVPGGTHITDAHGLMVLPGGVVHTRIQ 77
Db 4 IIKNGTIVTADTYEADLLIKDKIAMIGQHL-EEKGAEVIDAKGCVYFPGGIDSHTHLD 62
QY 78 MPVLGMPADDFCQCTKAALAGTTMILDHVPDPTGVSLAAYEQRERADSAACCDYSL 137
Db 63 MPFGTGTVDKDDFESCTIAAAGGTTIIDFCLTNKGEPLKKAETWHNKAKGKAVIDYGF 122
QY 138 HVDITRWHESKEELEALVKEGVNSFLVFMAYKDRCCSDSQMSYEIFSIIRDLGALAQV 197
Db 123 HLMISEITDDVLEELPKVIAEGITSEKVFMAKYNVFOADDGTLRTLVAAKELGALVMV 182
QY 198 HAENGDIIVEEQKRLLELIGITGPEGHVLSHPEEVEAEAVYRAVTIAKOANGLYVTKYMS 257
Db 183 HAENGDIVIDLTKKALAEAGNTEPIYHALTRPEVEGEATGRACQTLTAGSOLYVHVTC 242
QY 258 KGAADATAQAKRRGVVFGPEITASLGTDGSHYWSKNWAKAAAFVTPSPVNPDPPTADHL 317
Db 243 QAAVEKTAQARNKGLDVMGETCPOYLVDQS-YLEKPDFEGAKYVWSPPLR-EKWHQEV 300
QY 318 TLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPGTTNGIERMSVMWEKCVASGKMD 376
Db 301 WNALKNGQLTGLSDQCSDFKQKELGRGDTKIPNGGPMVEDRVSLFSEGVKKGRIT 360
QY 377 ENFVAVTSTNAAKIFNFYPRKGRVAVGSDADLVWNPATKIISAKTHNLNVEYNIFEG 436
Db 361 LNQFVDMSTRIAKLFLGFPKGTIAVGSADLVIFDPDIERSVISAETHHMAVDYNAFEG 420
QY 437 VECRGAUVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 421 MKVTGEPVSLCRGEFVVRDKQFVGKPGGYLKAGCF 458

RESULT 3

US-07-732-242C-3
; Sequence 3, Application US/07732242C
; Patent No. 5298399
; GENERAL INFORMATION:
; APPLICANT: Uozumi, Takeshi; Masaki, Haruhiko;
; APPLICANT: Hidaka, Makoto; Nakamura, Akira;
; APPLICANT: Maeda, Michihisa; Yoneta, Yasuo
; TITLE OF INVENTION: Gene of Urease
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frischauf, Holtz, Goodman & Woodward, P.C.
; STREET: 600 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016-2088
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.72mb
; COMPUTER: IBM PC compatible (NEC PC-9801 RX)
; OPERATING SYSTEM: MS DOS
; SOFTWARE: ASCII Form
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/732,242C
; FILING DATE: 19910718
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JPN 2-210178
; FILING DATE: 10-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Herbert
; REGISTRATION NUMBER: 17081

REFERENCE/DOCKET NUMBER: 910532/HG
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)972-1400
 TELEFAX: (212)370-1622
 TELEFAX: 236268
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 568 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-07-732-242C-3

Query Match 6.6%; Score 196; DB 1; Length 568;
 Best Local Similarity 19.8%; Pred. No. 1.2e-11;
 Matches 122; Conservative 83; Mismatches 211; Indels 200; Gaps 26;
 QY 1 MSFOGKSI-----PRITSDR---LLIRGRIVNDQSFYADVHVVEDGLIKOIGE--- 47
 Db 43 VFGGKVRIDGMCQHPLATSDVCLVLTNALIVDTYGIKADIGIKDGMISIGKAGN 102
 QY 48 -----NLIVPGGIHTIDAGHLMVPGGVVHTRLOMPVLGWTTPADDFCQGTAAALAGG 100
 Db 103 PLLMDGVDMVIGAAETVIAAGSMIVTAGGIDAHIFICP-----QQIETALASG 151
 QY 101 -TTIILHVFDPDTGVSLLAAEQWRERADSAACDYSLVHVDITRWHSIKEELEALVKEK 159
 Db 152 VTTMIGGTGTGATG---NATTC-----TPGPNHRLQA----- 184
 QY 160 GVSFLFVMAVMDRCQCSQMYEIFSIIRDLGALAAQVHAENGDIIVEEQRRLLELGITG 219
 Db 185 -AEEFNLGFLGKNCSD-----EAPLKEQIEAGAVG 216
 QY 220 PEGHVLSHPEE--VBAEAVYRAVTIAKANCPLYV-TKVMKG--AADAIAQAKRRGVVV 274
 Db 217 LKLH-----EDWGSTAAIDTCLKVADRYDVQVAIHTDLNEGGVETLKAIDGRVHT 271
 QY 275 FCEPITASLGTGSHYWKAKAAAFVTSPPVNDP-----TTADHLTCLLSSGDLQ 327
 Db 272 Y-----HTEGAGGH--APDIIKAAGFPNLPSTNTPRYTINTLEHDLMLVCHHLD 324
 QY 328 VTGSAHCTTTAKAVKDN-----FALIPETNG---IEERMSVWKEC--- 369
 Db 325 ANIPEDIAF--ADSRIRKETIAAEDVLHDLGVFSMISSQAMGRVGEVITITWTADKM 382
 QY 370 -VASGKMDEN-----EFVAVTSTNAKIFNFPKRGVAVGSDADLVINPKATK 418
 Db 383 KKGKGLQEDNGVDNFRVKRYIAKTYINPAIAGHTADYVGSVEGKLADLVVNP--- 438
 QY 419 IISAKTHNLNVEYNIEFEGVECRGAPAVVISQGRVALED-----CKMFVT 462
 Db 439 -----AFFGVK---PELVKGGMIAYSTMGDPNASIPTPPVLYRPMFAA 480
 QY 463 PGAGR-----FVPRFTFPDFVYKRIKARNRLAEIHGV-----PRGLYDGP 502
 Db 481 KGDARYQTSTTFVSKAAEYKGIHQGLKKKVPVHGIRKLYKKDLLNDKTPKIDVDQ 540
 QY 503 VHEVMVPKPGSGAPA 518
 Db 541 TYEVKVDGQLVTCPEA 556

RESULT 4
 US-08-967-513-5
 : Sequence 5, Application US/08967513
 : Patent No. 5783436
 : GENERAL INFORMATION:
 : APPLICANT: Robert P. Hausinger
 : TITLE OF INVENTION: Mutant Urease and Method
 : TITLE OF INVENTION: of Use For Determination
 : TITLE OF INVENTION: of Urea

NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ian C. McLeod
 STREET: 2190 Commons Parkway
 CITY: Okemos
 STATE: Michigan
 COUNTRY: USA
 ZIP: 48864
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/967,513
 FILING DATE: 11-NOV-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/687,645
 FILING DATE: July 26, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ian C. McLeod
 REGISTRATION NUMBER: 20,931
 REFERENCE/DOCKET NUMBER: MSU 4.1-309
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (517) 347-4100
 TELEFAX: (517) 347-4103
 TELEX: NO. 5783436e
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 567
 TYPE: amino acids
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 MOLECULE TYPE:
 DESCRIPTION: protein
 HYPOTHETICAL: No
 ANTI-SENSE: No
 ORIGINAL SOURCE:
 ORGANISM: N/A
 STRAIN: N/A
 INDIVIDUAL ISOLATE: N/A
 CELL TYPE: N/A
 FEATURE:
 NAME/KEY: subunit UreC
 LOCATION:
 IDENTIFICATION METHOD: Sequencing
 OTHER INFORMATION: Encoded subunit of mutant
 OTHER INFORMATION: urease
 US-08-967-513-5

Query Match 4.8%; Score 142; DB 1; Length 567;
 Best Local Similarity 20.8%; Pred. No. 5.9e-06;
 Matches 122; Conservative 69; Mismatches 212; Indels 184; Gaps 24;
 QY 17 LLIRGRIVNDQSFYADVHVVEDGLIKOIGE-----NLIVPGGIHT--IDAHLGLV 66
 Db 68 LVLTNALIVDHGIVKADIGVKGDRIFAGKAGNPDIOPNVTIPIGAATEVIAEGKIYT 127
 QY 67 PGVDVHTRLOMPVLGWTTPADDFCQGTAAALAGGTTMILHDVFPDGTGVSLLAAEQWRER 126
 Db 128 AGGIDTHIHWICPQ-----QAEALVSGVITMVGGGTGPAAGT----- 165
 QY 127 ADSAACCDYSLHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCSSQMYEIFS 186
 Db 166 -HATTCFPG-----PWYIS-----RMLQAAD-----S 186
 QY 187 IIRDLGALAAQVHAENGDIIVEEQRRLLELGITGPEGHVLSHPEEVA--EAVYRAVTIAK 244
 Db 187 LPVNIIGLKGNSQPDALREQ-----VAAGVIG-----LKIQEDWGATPAADICALTVD 237

```

QY 245 QANCPLYV---TKVMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNWKAAAF 301
Db 238 EMDIQVALHSDTLNESGFVEDTLAAIGGRTIHTF-----HTEGAGGGH--APDIITACAH 290
QY 302 VTSPPVNPDPPTA-----DHLICLLSSDGLQVTVGSAHCTF-----TTAQKAVGKD 346
Db 291 PNILPSSNTPLPYTLNTIDHLDMLVCHHLDPPDIAEDVAFASRRIRRETIAAEDVLHD 350
QY 347 --NFALIPEGTNG---IEERMSVMWEKC---VAGKMDKDE-----NEFVAVTSTN 387
Db 351 LGAFSLTSSDQAMGRVGEVILRTVOVAHRMKVQRGALAEETGDNDFRVKRYIAKYTN 410
QY 388 AAKIFNFPKRGVAVGSDADLVINPKATKIISAKTHNLNVEYNIIEGVECRGAPAVVI 447
Db 411 PALTHGIAHEVGSIEVGLADLVVWSP-----AFEGVK----PATVI 448
QY 448 SQGRVALED-----GKMFVTPGAGRFVPRKTF-----PDFVYKRIKAR 485
Db 449 KGMIAIAPMGDINASITPTQPVHYRPMFGALGSARHCHRLTFLSQAAAANGVAERLNL 508
QY 486 NLAIEIHG-----PRGLYDGPVHEVMVPAKPGSGAPA 518
Db 509 SAIAVVKGCRTVOKADMVHNSLPQNTITVDAQTYEVRVDGELITSEPA 555

```

RESULT 5

US-08-687-645B-5

; Sequence 5, Application US/08687645B

; Patent No. 5846752

; GENERAL INFORMATION:

; APPLICANT: Robert P. Hausinger

; TITLE OF INVENTION: Mutant Urease and Method

; TITLE OF INVENTION: of Use For Determination

; TITLE OF INVENTION: of Urea

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ian C. McLeod

; STREET: 2190 Commons Parkway

; CITY: Okemos

; STATE: Michigan

; COUNTRY: USA

; ZIP: 48864

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/687.645B

; FILING DATE: July 26, 1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Ian C. McLeod

; REGISTRATION NUMBER: 20,931

; REFERENCE/DOCKET NUMBER: MSU 4.1-309

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (517) 347-4100

; TELEFAX: (517) 347-4103

; TELEX: No. 5846752e

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 567

; TYPE: amino acids

; STRANDEDNESS: Single

; TOPOLOGY: Linear

; MOLECULE TYPE:

; DESCRIPTION: protein

; HYPOTHETICAL: No

; ANTI-SENSE: No

```

; ORIGINAL SOURCE:
; ORGANISM: N/A
; STRAIN: N/A
; INDIVIDUAL ISOLATE: N/A
; CELL TYPE: N/A
; FEATURE:
; NAME/KEY: subunit UreC
; LOCATION:
; IDENTIFICATION METHOD: Sequencing
; OTHER INFORMATION: Encoded subunit of mutant
; OTHER INFORMATION: urease
; US-08-687-645B-5

```

```

Query Match 4.8%; Score 142; DB 2; Length 567;
Best Local Similarity 20.8%; Pred. No. 5.9e-06;
Matches 122; Conservative 69; Mismatches 212; Indels 184; Gaps 24;

```

```

QY 17 LLIRGGRIYNDQSFYADVHVHVEDGLIKOIGE-----NLIVPGGIHT--IDAHLMLV 66
Db 68 LVTNALIVDHGIVKADIGVKDGRIFAIGKAGNPDIQPNVTIPIGAATEVIAEGKIYT 127
QY 67 PGVDVHTRLQMPVLGMPADDFCQGTKAALAGCTTMLDHFVFDPTGVSLAAYEQWRER 126
Db 128 AGGIDTHIHICFQ-----QAEELVSGVTTVMVGGGTGPAAGT----- 165
QY 127 ADSAACCDYSLHVDITRWHSIKEELEALVKEGVNSFLVFMAYKDRCCQSDSQMYEIFS 186
Db 166 --HATTCTPG-----PWYIS-----RMLQAD-----S 186
QY 187 IIRDLGALQVHAENGDIVEEQKRLLELITGPEGHVLSHPEVEA--EAVYRAVTIAK 244
Db 187 LPVNIIGLLKGNVSQPDALREQ---VAAGVIG---LKIQEDWGATPAIDCALTVAD 237
QY 245 QANCPLYV---TKVMSKGAADAIAQARRGVVVFGEPIITASLGTDGSHVSKNWKAAAF 301
Db 238 EMDIQVALHSDTLNESGFVEDTLAAIGGRTIHTF-----HTEGAGGGH--APDIITACAH 290
QY 302 VTSPPVNPDPPTA-----DHLICLLSSDGLQVTVGSAHCTF-----TTAQKAVGKD 346
Db 291 PNILPSSNTPLPYTLNTIDHLDMLVCHHLDPPDIAEDVAFASRRIRRETIAAEDVLHD 350
QY 347 --NFALIPEGTNG---IEERMSVMWEKC---VAGKMDKDE-----NEFVAVTSTN 387
Db 351 LGAFSLTSSDQAMGRVGEVILRTVOVAHRMKVQRGALAEETGDNDFRVKRYIAKYTN 410
QY 388 AAKIFNFPKRGVAVGSDADLVINPKATKIISAKTHNLNVEYNIIEGVECRGAPAVVI 447
Db 411 PALTHGIAHEVGSIEVGLADLVVWSP-----AFEGVK----PATVI 448
QY 448 SQGRVALED-----GKMFVTPGAGRFVPRKTF-----PDFVYKRIKAR 485
Db 449 KGMIAIAPMGDINASITPTQPVHYRPMFGALGSARHCHRLTFLSQAAAANGVAERLNL 508
QY 486 NLAIEIHG-----PRGLYDGPVHEVMVPAKPGSGAPA 518
Db 509 SAIAVVKGCRTVOKADMVHNSLPQNTITVDAQTYEVRVDGELITSEPA 555

```

RESULT 6

US-08-920-095-3

; Sequence 3, Application US/08920095

; Patent No. 5837240

; GENERAL INFORMATION:

; APPLICANT: Cynthia K. Lee et al.

; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT UREASE VACCINE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

Tue Jul 31 13:08:17 2001

```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920.095
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/431.041
; FILING DATE: 28-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-920-095-3

```

```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920.095
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/431.041
; FILING DATE: 28-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/013001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-920-095-3

```

```

Query Match 4.6%; Score 136.5; DB 2; Length 566;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 121; Conservative 80; Mismatches 220; Indels 211; Gaps 29;

QY 17 LLIRGRVNDQSFYADVHEDGLIKQGE-----NLIVPGGIHTIDAHGLM 64
DB 68 LIITNALIVDTGIYKADIGIKDGKTAGIKGKNDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVVDVHTRLOMPVLGMPADDFCQGTAAALAGG--TTMILDHVPDPGVSLAAEQW 123
DB 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGTGPGADGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIEKELEALVKEGVNSFLVFMAYKDRCCSDSOMYE 183
DB 170 -----TTTPGRRNLKWLRA-AEEYSMN----- 192
QY 184 IFSIIRDLGALAQVHAENGDIIVEEQKRLLELGITG-----PE--GHVLSHPPEV 231
DB 193 -----LGFLAKGNASNDASLADQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRVTIAKQANCPLYVTKVMSKGAADAIQAQKRRGVVVFGEPIITASLGTDGSHYW 291
DB 242 DVQVAIHTDTL-NEAGC-----VEDTMAAIAGRTHWTF-----HTEGAGGGH-- 282
QY 292 SKNKAFAAFVTSPPVNPDP-----TTADHLTCLSSGDLQVYVTSAGHCTFTTA-----Q 340
DB 283 APDIIVAGEHNILPASTNPTPTPTVNTAEHMDMLVCHLDKSIKEDQVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPGTNG--IEERMSWVEKCVAS-----GKMDENE----- 379
DB 343 TIAAEDTLHDMGIFSTSSDQAMGRVGEVITRTWTADKNKEFGRLKEEGKDNDFRI 402
QY 380 --FVAVTSTNAKIFNYPKRGVAVGSDADLVINWPKATKIISAKTHNLNVEYVIFEGV 437
DB 403 KRLSKYTIINPAHTAHGISEVGSVEGKVADLVLWSP-----AFFGV 444
QY 438 ECRGAPAVVISQGRVALEDKGMFTVTCAGRFVPRKFTFPFVYKRI-----KARNELAEI 491
DB 445 K-----PNMIKGGFIALSQ-----MGDANASIPT---PQVYVYREMAFHGKAKYD-ANI 491
QY 492 HGVPRGLYDGPVHEVM-----VPAPGSGAPARASCPGKISVPPVRNLHQSGFSLSGSO 545
DB 493 TVFQAAYDKGIELGELGRQVLVVK-----NC-----RNITKMDQFNDDTT 533

```

```

QY 546 ADDEIARKT-----AQKIMAPPGGRSNITSL 571
DB 534 AHIEVNPETYHVFDGKEVTSKPANKVSLAQL 565

RESULT 7
PCT-US96-05800-3
; Sequence 3, Application PC/TUS9605800
; GENERAL INFORMATION:
; APPLICANT: Oravax, Inc.
; TITLE OF INVENTION: MULTIMERIC, RECOMBINANT DREASE VACCINE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05800
; FILING DATE: 23-APR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,041
; FILING DATE: 28-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/568,122
; FILING DATE: 06-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-05800-3

```

```

Query Match 4.6%; Score 136.5; DB 5; Length 566;
Best Local Similarity 19.1%; Pred. No. 2.2e-05;
Matches 121; Conservative 80; Mismatches 220; Indels 211; Gaps 29;

QY 17 LLIRGRVNDQSFYADVHEDGLIKQGE-----NLIVPGGIHTIDAHGLM 64
DB 68 LIITNALIVDTGIYKADIGIKDGKTAGIKGKNDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVVDVHTRLOMPVLGMPADDFCQGTAAALAGG--TTMILDHVPDPGVSLAAEQW 123
DB 128 VTAGGIDTHIFISP-----QQIPTAFASGVTTMIGGTGPGADGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIEKELEALVKEGVNSFLVFMAYKDRCCSDSOMYE 183
DB 170 -----TTTPGRRNLKWLRA-AEEYSMN----- 192
QY 184 IFSIIRDLGALAQVHAENGDIIVEEQKRLLELGITG-----PE--GHVLSHPPEV 231
DB 193 -----LGFLAKGNASNDASLADQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRVTIAKQANCPLYVTKVMSKGAADAIQAQKRRGVVVFGEPIITASLGTDGSHYW 291

```

Db 242 DVQVAHTDTL-NEAC-----VEDTMAAIACTMTF-----HTEGAGGHH-- 282
QY 292 SKNKAFAAVTSPVNPDP-----TTADHLTCLSSGDLQVTSACHTFTA-----Q 340
Db 283 APDIKVAEHNLPASTNPTPTVTNTAEHMDMLVCHHLKSIKEDVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPGTNG-----IERMMSWWEKCVAS-----GKMDENE----- 379
Db 343 TAAEDTLHDMGIFSTSSDSQAMGRVGEVITRTWQADKNKKEGRLEKEEGKNDNFR 402
QY 380 --FVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKATKIIISAKTHNLNVEYNIEFV 437
Db 403 KRYLSKYTINPAIHAGISEYVGSVEGVKADLVLMSP-----AFFGV 444
QY 438 ECRGAPAVVISQGRVLEBGMFVTPGAGRVFPRTKTFDFVYKRI-----KARNLAEL 491
Db 445 K-----PNMIKGGFTALSO-----MGDANASIPT---PQPVYIREMAHGHKAKYD-ANI 491
QY 492 HGVPRLGDLGVPHEVM-----VPAKPGSGAPARASCPGKISVPPVRLNHLQSGFSLSGSQ 545
Db 492 TFSVQAAYDKIGELGLERQVLPK-----NC-----RNITKDKMQFNDDTT 533
QY 546 ADDHIART-----AOKIMAPPGGRSNTISL 571
Db 534 AHIEVNPETHYFVFDGKEVTSKPAKVSLAQL 565

RESULT 8

US-08-467-822-21
; Sequence 21, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 569 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..569
; OTHER INFORMATION: /note= "URE B - FIGURE 3."
US-08-467-822-21

Query Match 4.6%; Score 136.5; DB 2; Length 569;
Best Local Similarity 19.7%; Pred. No. 2.3e-05;
Matches 109; Conservative 65; Mismatches 198; Indels 181; Gaps 24;

QY 1 MSFOGKKS-----PRITSDRLLIRGGRIVNDQSFYADVHVEDGLIKQICE--- 47
Db 43 IKFGGKTIIRDGMSQTSNPSSEYELDLVLNLIYDVTGIYKADIGIKGKIAGIKRAGN 102
QY 48 -----NLIVPGGIHTIDAHGLVLPFGVDVHTRLOMPVLGMPADDFCOGTKAALAG 99
Db 103 DMQDGVNNLCVGPATEALAAEGLIVTAGDITHIFISP-----QOIPTAFAS 151
QY 100 G-TTMILDHVPDPTGVSLAAYEONRERADSAACDYSLSHVDITRWHESIKEELEALYKE 158
Db 152 GVTTMIGGTGPDAGTNA-----TTITPGRANLKSMURA-AEE 188
QY 159 KGVNSFLVFMAYKRCQSDSQMEI-----FSIIRDLGAL-AQVHAENGDIVEEQKR 211
Db 189 YAMN--LGFLA-KGNVSEPSLRDOIEAGIGFKITHDWGSTPAAIH----- 232
QY 212 LLELGITPEGHVLSHPEVEAEAVYRAVTIAKQANCPLYVTVKMSKGAADAIQAARRG 271
Db 233 -----HCLNVADEYDVQVAIHTDTL-NEAGC-----VEDTLEAIAGR 269
QY 272 VVTFGEPTASLGTDGSHYNSKNWAKAAAFVTSPPVNDP-----TTADHLTCLSSG 324
Db 270 IHTF-----HTEGAGGGH--APDIKMAEFNLPASTNPTPTKNTAEHMDLAVCH 322
QY 325 DLQVTSACHTFTA-----QKAVGKDN-----FALIPGTNG-----IERMMSWWEKCV 371
Db 323 HLDSIKEDYQFADSRIRPQTIAEDQLHDMGIFSTSSDSQAMGRVGEVITRTWQADK 382
QY 372 S-----GKMDENE-----FVAVTSTNAKIFNFPYPRKRVAVGSDADLVINPKAT 417
Db 383 NKKEFGRLKEEGKNDNFRISKYITNPGIAHGISDYVGSVEGVKADLVLMSP--- 439
QY 418 KIISAKTHNLNVEYNIEFVGEVGRGAPAVVISQGRVALED-----GKMEV 461
Db 440 -----AFFGIK-----PNMIKGGFTALSO----- 480
QY 462 TPCAGRFVPRKTF 474
Db 481 HGGKPKFDNTIF 493

RESULT 9

US-08-432-697-21
; Sequence 21, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thiberge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; POLYPEPTIDES

Db 68 LIITNALIVDYTGKADIGKGIAGKGNKMDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVDVHTRLOMPVLGTMTPADDFCQGTKAALAGG-TTMIIDHVPFDTGVSLLAAYEQW 123
Db 128 VTAGGIDTHIFISP-----QQPTAFASGVTTMIGGGTGPDAGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIKEELEALVKKGVSFLVFWAYKDRQCSDSQMYE 183
Db 170 -----TTITPGRNKLWMLRA-AEYSNN----- 192
QY 184 IFSIIRDLGALAOVHAENGDIVEEQKRLLELGITG-----PE--GHVLSHPPEV 231
Db 193 -----LGFLAKGNASN-DASARDQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRAVTTAKOANCPLYYTKVMSKGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYW 291
Db 242 DVQVAIHDTL-NEAGC-----VEDTMAAIAGRTHMTF-----HTEGAGGHH-- 282
QY 292 SKNWAACAAFTVSPVNDP-----TTADHLTCLSSGDLQVTSAGHCTFTA-----Q 340
Db 283 APDIKIVAGEHNILPASTNPTFTVTNTEAHEMDMLVCHHLDKSIKEDVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPETNG-----IEERMSVMWKCVA-----GKMDENE----- 379
Db 343 TIAAEDTLHDMGIFESITSSDSQAMGRVGEVITRTWQTADANKKEFGKKEEGDNDNFRI 402
QY 380 --FVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKNKATKIISAKTHNLNVEYNIFEGV 437
Db 403 KRYLSKTYINPAIAHGISVGVSEVGVKADVLWSP-----AFFGV 444
QY 438 ECRGAPAVISQGRVALEDGKMEVTPCAGRFVPRKFTFPDVPYKRIKARNLAEIHG----- 493
Db 445 K-----PNMIKGGFIALSQ-----MGDANASIPT---POPVIYR-----EMFGHGKAKY 487
QY 494 -----VPRGLYDGPVHEVM-----VPAKPGSGAPARASCPCGKISVPPVRLNHSOGFSL 541
Db 488 DRNITFVSQAAYDKIGKELGLERQVLPVK-----NC-----RNITKDMQF 529
QY 542 SGSOADHDHART-----AQKIMAPPGGRSNTLSL 571
Db 530 NDTTAHIEVNPETYHYFVDGKEVTSKPANKVSLAQL 565

RESULT 11

US-08-432-697-26
; Sequence 26, Application US/08432697
; Patent No. 6248330
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/432,697
; FILING DATE: 02-MAY-1995

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-697-26

Query Match 4.5%; Score 134.5; DB 4; Length 569;
Best Local Similarity 19.0%; Pred. No. 3.7e-05;
Matches 121; Conservative 78; Mismatches 218; Indels 219; Gaps 30;
QY 17 LLIRGRIYNDQSFYADVHVHVEDGLIKQIGE-----NLIYVPGIHTIDAHGLM 64
Db 68 LIITNALIVDYTGKADIGKGIAGKGNKMDQGVKNLSVGPATEALAGEGLI 127
QY 65 VLPGGVDVHTRLOMPVLGTMTPADDFCQGTKAALAGG-TTMIIDHVPFDTGVSLLAAYEQW 123
Db 128 VTAGGIDTHIFISP-----QQPTAFASGVTTMIGGGTGPDAGTNA----- 169
QY 124 RERADSAACCDYSLHVDITRWHSIKEELEALVKKGVSFLVFWAYKDRQCSDSQMYE 183
Db 170 -----TTITPGRNKLWMLRA-AEYSNN----- 192
QY 184 IFSIIRDLGALAOVHAENGDIVEEQKRLLELGITG-----PE--GHVLSHPPEV 231
Db 193 -----LGFLAKGNASN-DASARDQ-----IEAGAIGFKIHEDWGTTPSAINHALDVADKY 241
QY 232 EAEAVYRAVTTAKOANCPLYYTKVMSKGAADAIAQAKRRGVVVFGEPIITASLGTDGSHYW 291
Db 242 DVQVAIHDTL-NEAGC-----VEDTMAAIAGRTHMTF-----HTEGAGGHH-- 282
QY 292 SKNWAACAAFTVSPVNDP-----TTADHLTCLSSGDLQVTSAGHCTFTA-----Q 340
Db 283 APDIKIVAGEHNILPASTNPTFTVTNTEAHEMDMLVCHHLDKSIKEDVQFADSRIRPQ 342
QY 341 KAVGKDN-----FALIPETNG-----IEERMSVMWKCVA-----GKMDENE----- 379
Db 343 TIAAEDTLHDMGIFESITSSDSQAMGRVGEVITRTWQTADANKKEFGKKEEGDNDNFRI 402
QY 380 --FVAVTSTNAKIFNFPYPRKGRVAVGSDADLVINPKNKATKIISAKTHNLNVEYNIFEGV 437
Db 403 KRYLSKTYINPAIAHGISVGVSEVGVKADVLWSP-----AFFGV 444
QY 438 ECRGAPAVISQGRVALEDGKMEVTPCAGRFVPRKFTFPDVPYKRIKARNLAEIHG----- 493
Db 445 K-----PNMIKGGFIALSQ-----MGDANASIPT---POPVIYR-----EMFGHGKAKY 487
QY 494 -----VPRGLYDGPVHEVM-----VPAKPGSGAPARASCPCGKISVPPVRLNHSOGFSL 541
Db 488 DRNITFVSQAAYDKIGKELGLERQVLPVK-----NC-----RNITKDMQF 529
QY 542 SGSOADHDHART-----AQKIMAPPGGRSNTLSL 571
Db 530 NDTTAHIEVNPETYHYFVDGKEVTSKPANKVSLAQL 565

RESULT 12

US-08-467-822-25
; Sequence 25, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes

APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 840 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-25

Query Match 4.3%; Score 129; DB 2; Length 840;
Best Local Similarity 19.3%; Pred. No. 0.00028;
Matches 101; Conservative 58; Mismatches 181; Indels 182; Gaps 20;

QY 15 DRLIRGRIVNDQS-----FYADVHVEDGLIKOIGE- 47
DB 311 DECVFGGKVRIDMGQSCGHPAISLDVITNAVIVDTYGIKADIGIKDGLIASIGKA 370
QY 48 -----NLIVFGGIHTIDAGLMLVPGVDVHTRLQMPVLGWTADDFCQGTAA 96
DB 371 GNPDMNGVFSNMIGANTEVIAGEGLIVTAGIDCHIHVICPOLVY-----EAI 420
QY 97 LAGTTHMLDHPDPTGVSLAAYEQWRERADSAACCDYSLHVDITRWHSIKEELEALV 156
DB 421 SSGITLVGGGTGPAAGT-----RATT-----442
QY 157 KEKGVNSFLVMAYKDRQCQSDSQMVEIFSIIIRDLAGAQAQVHAENGDIIVEEQRLLELG 216
DB 443 -----CTSPQOMRLMLQSTDLLPLNFGFTGKSSSKPDELHEIIGAG 485
QY 217 ITGPEGHVLSPHEE--VEAEAVYRANTIAKQANCPYV--TKVMSKG--AATAQAQRKG 271
DB 486 AMGLKLH-----EDWGSTPAIDNCLTIAEHHDITQIINHDTLTINEAGFVHSHIAAFKGR 540

QY 272 VVVFGEPIASLGTGSHYWSKNWAKAAAFVTSPPVNPDPPTA-----DHLTCLLSG 324
DB 541 IHVY-----HSEGAGGH--APDIKVCCKNVLPSTNTRPLTSITDEHLDMLWVCH 593
QY 325 DL-----QVTSAGHCTP---TTAQKAVGKDNFALI-----PEGTNGTEERMSMWVE---- 367
DB 594 HLDREIPEDVAFHSRIRKKTIAAEDVLHDIGAIISSDSQAMGRVGEVISRTWOTADK 653
QY 368 -----KCVASGKMDENEF-----VAVTSTNAKIFNFPKGRVAVGSDADLVWNP 414
DB 654 NKAQTGPLKCDSS---DNDNFRIKRYIAKTYINPAIAHGISOYVYGSVEGKLADLVLMKP 710
QY 415 KATLIISAHTNLNVEYNEFEGVCEGCPAPVVISOGRAVED 456
DB 711 S-----FFGTK-----PEWIKGGHVAWAD 730

RESULT 13
US-08-432-697-25
Sequence 25 Application US/08432697
Patent No. 6248330
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauerbaum, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0137-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 840 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-432-697-25

Query Match 4.3%; Score 129; DB 4; Length 840;
Best Local Similarity 19.3%; Pred. No. 0.00028;
Matches 101; Conservative 58; Mismatches 181; Indels 182; Gaps 20;

QY 15 DRLIRGRIVNDQS-----FYADVHVEDGLIKOIGE- 47
DB 311 DECVFGGKVRIDMGQSCGHPAISLDVITNAVIVDTYGIKADIGIKDGLIASIGKA 370

QY 48 -----NLVPGGIHTIDAHGLMVLPGGVVDVHTRLOMPVLGMPADDFCQGTAA 96
Db 371 GNPDIMNGVSNMIGANTVIAAGELVITAGDICHIIYICPOLVY-----EAI 420
QY 97 LAGGTTMILDHVPDPTGVSLLAAAYEOWRERADSAACCDYSLHVDITRWHESIKEELEALV 156
Db 421 SSGITTLVGGGTGPAAGT-----RATT----- 442
QY 157 KEGVNSFLVMAYKDRCCSDSOMYEIISIRDLGALAQAHAENGDIIVEBOKRLLLELG 216
Db 443 -----CTPSPQMLMLQSTDDLPNFGFTGKSSSKPDELHEIKAG 485
QY 217 ITGPEGHVLSHPEE--VEAEAVYAVTIAKQANCPLYV-TKVMKSG--AADATAQAARRG 271
Db 486 AMGLKHL-----EDWGSTPAADNCLTIAEHHDIOINIHTDITLNEAGFVHSHIAAFKGR 540
QY 272 VVVEGEPITASLTGDSHWSKWKAAAFVSPVNPDPPTA-----DHLTCLSSG 324
Db 541 IHTY-----HSEGAGGH--APDIKVCIGKNVLPSTNPTPLTSNTIDEHLDMLMVCH 593
QY 325 DL-----QVTGSAHCTF--TTAAKAVGKDNFALI-----PECTNGIEERMSMWVE--- 367
Db 594 HLDRETPDVAFAHSIRKKTIAEDVLHDIGAISIISDSQAMGRVGEVISRTWTOTADK 653
QY 368 -----KCVASGKMDENEF-----VAVTSTNAKIFNFPKGRVAVGSDADLVINWP 414
Db 654 NKAQTGPLKCDSS--DNDNFRKRYIAKTYINPAIAHGISQYVGSVEVGLADLVWKP 710
QY 415 KATKIISAKTHNLNVEYNIFEGVEGRCAPAVVISQGRVALED 456
Db 711 S-----FFGTK-----PEMVIKGMVAWAD 730

RESULT 14
5206163-3
; Patent No. 5206163
; APPLICANT: RENARD, ANDRE; DINA, DINO; MARTIAL, JOSEPH
; TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
; VIRUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 331,037
; FILING DATE: 29-MAR-1989
; APPLICATION NUMBER: 752,981
; FILING DATE: 08-JUL-1985
; SEQ ID NO:3
; LENGTH: 2616
5206163-3

Query Match 3.8%; Score 112.5; DB 6; Length 2616;
Best Local Similarity 20.7%; Pred. No. 0.12; Indels 99; Gaps 13;
Matches 72; Conservative 45; Mismatches 131; Indels 99; Gaps 13;

QY 125 ERASAAACCDYSLHVDITRWHESIK-----EELEALVKEKGVNSFLVEMAYKDRCCQ 176
Db 2075 ENQOSPGLHDLKLEIFHTIAQPSLKHVTGYETWQLEAGINRKGAGFL---EKNLGEV 2131
QY 177 SDSOMYEIISIRDLGALAQAHA-----ENGDIVEEEOKRLLLELGITG 219
Db 2132 LDSEKHLVDIIRLTKGRKIRYETAITPKNEKRDVSDMQAGDIVDEKKPR----- 2183
QY 220 PEGHVLSPHEVEAEV--YRAVTIAKQANCPLYVTKVMKSGAADAATAQAARRGVVVEGE 277
Db 2184 ----VIQYPEAKTLATIKYMNWVKQ---PVVIPGEGKTPLEKIFNKVRKEWDLFNE 2236
QY 278 PITASLTGDC-----SHWSKWKAAAFVSPVNPDPPTADHLTCL 320
Db 2237 PVAVSFDTKAWDTQVTRDLRLIGEIQYRYKEWHKFIDITIDHVMVEVPVITAD----- 2291

QY 321 LSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPECTNGIEERMSVMVKECVASGKMDENEF 380
Db 2292 ---GEVYI-----RNGQSGSQPD-----TSAGNSMLNVLTMWYAFCESTGV----- 2330
QY 381 VAVTSTNAKIFNFPKGRVAV-GSDADLVINWPKATKIISAKTHN 426
Db 2331 -----PYKSFN---RVARIHVGDDGDFLITERGLGTGKICQORDAN 2367

RESULT 15
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN USA
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1611 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-5

Query Match 3.4%; Score 102.5; DB 2; Length 1611;
Best Local Similarity 21.5%; Pred. No. 0.56;
Matches 129; Conservative 73; Mismatches 236; Indels 161; Gaps 30;

QY 33 ADVHVEDGLIKQI-----GENLVPGGIHTIDAHGLMVLPGGVVDVHTRLOMPVLGMPADDD 88
Db 691 AVVALRAGLIRYLAGRAMAVALPAGEVEA-GLAKWP-GVEV-----AAVNG--PAST 741
QY 89 FCQGTKAALAGGTTM-----ILDHVFPDPTGVSLAAAYEQWRERADSAACCDYSLHVDITR 143
Db 742 VVSGDRRAVAVVAVCAQEGVQARLIP-----VDYASH---SR 776
QY 144 WHESIKEELEALVKEKGVNSFLVEMAYKDRCCSDSOMYEIISIRDLGALAQAHAENGND 203
Db 777 HVEDLKELELVLSGIRSPRPVPCSTVAGEQCEPFDAGYFNRLNRNVEFSAVVVG 836
QY 204 IVEEEOKRLLLELGITGPEGH-VLSHPEVEAEAVYRAV---TIKQANCPLYVTKVMK 258
Db 837 LLEBGRHREFIEV-----SAHPVLVHALQETAADRSVHATGTLRRQDDSP-----H 883
QY 259 GAADIAQAARRGVVVEGEPI-----TASLGT-----DGSYWSKWKAAAFVTPPVN-- 308

```
Db 884 RLLTSTAFAWAGATLTWDPALPPGHLTTLPTYPFNHHYWLDTPTTPTTQSPDADW 943
QY 309 -----PDPTTADHLTCLL-----SSGDLQVTGSAHCTTTAAKA 342
Db 944 RYRVTKALTEESTPASSPSGHLLVTPPTPEGRTLGDRAGALARQCATVERLVVDPVA 1003
QY 343 VGKDNFALIEGTNGIEERMSMVKEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVA 402
Db 1004 VGRD-----GLAARLGERWDGVLSSLGADERPLRHPALNRAVMGTLLAQAAALD 1053
QY 403 VGSDADLVINPKATKIISAKTHNLNVEYNIEFEGVECRGAPAVVISQGR-VALEDGKMFV 461
Db 1054 AGCEARINAVTREAVAVSPSEVP-----RDAGQLWGLGRGIALHPSLW- 1098
QY 462 TPCAGRFVPRKTFPDPFVYKRIKAR--NRLAEIHGVPRLYDGPVHEVMVPAPKPGSGAPAR 519
Db 1099 ----GGLIDLPAVPD---ERAWARAVRRL-----VPHG-----EDQIAAR-ASGAYGR 1138
QY 520 ASCPGKISVPPV---RNLHQSGFSL--SGSQA-DDHIARRTAQKIMAPPGGRSNITSLS 572
Db 1139 RLLP-----APPAASRRTCTPSGTVLVTGGTGALGGHLARRLAR-----GGTGHLVLTS 1187
```

Search completed: July 30, 2001, 11:40:58
Job time: 282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:35:11 ; Search time 21.29 Seconds

(without alignments)
1628.789 Million cell updates/sec

Title: US-09-367-496-8

Perfect score: 2985

Sequence: 1 MSFQKKSIPRTSDRLIR.....RTAQKIMAPPGCRSNTLSLS 572

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2882	96.5	553	19 AAW68489	Human partial ULIP
2	2803	93.9	572	19 AAW68488	Mouse ULIP-4 prote
3	2325	77.9	572	19 AAW68486	Mouse ULIP-1 prote
4	2151	72.1	572	19 AAW68487	Mouse ULIP-3 prote
5	1044	35.0	683	21 AAG46568	Arabidopsis thalia
6	1011.5	33.9	485	17 AAW03025	Pseudomonas hydan
7	945	31.7	588	21 AAG46569	Arabidopsis thalia
8	901	30.2	576	21 AAG46570	Arabidopsis thalia
9	837	28.0	471	13 AAR29715	Heat stable hydan
10	826	27.7	460	16 AAR69874	B.Chermoglucosid
11	691.5	23.2	457	16 AAR82837	Agrobacterium radi

12	684.5	22.9	457	17 AAW03023	Bacillus sp. hydan
13	684.5	22.9	457	17 AAW03024	Agrobacterium hyda
14	600	20.1	167	21 AAB58948	Breast and ovarian
15	534	17.9	281	21 AAG23637	Arabidopsis thalia
16	464	15.5	283	21 AAG23636	Arabidopsis thalia
17	464	15.5	284	21 AAG23635	Arabidopsis thalia
18	408	13.7	153	21 AAB41399	Human ORFX ORF1163
19	396.5	13.3	458	21 AAB26143	Arthrobacter hydan
20	396.5	13.3	458	21 AAB26146	Arthrobacter hydan
21	395.5	13.2	458	21 AAB26141	Arthrobacter hydan
22	394.5	13.2	458	21 AAB26140	Arthrobacter hydan
23	394.5	13.2	458	21 AAB26144	Arthrobacter hydan
24	394.5	13.2	458	21 AAB26145	Arthrobacter hydan
25	390.5	13.1	458	21 AAB26142	Arthrobacter hydan
26	242	8.4	447	22 AAB80122	Corynebacterium g1
27	240	8.0	506	21 AAG26653	Arabidopsis thalia
28	240	8.0	512	21 AAG26652	Arabidopsis thalia
29	216	7.2	397	22 AAB50541	B. lactofermentum
30	206	6.9	409	21 AAG26654	Arabidopsis thalia
31	200.5	6.7	129	21 AAB40494	Human ORFX ORF258
32	193	6.5	568	14 AAR36387	Urease gamma subun
33	178.5	6.0	169	21 AAB43021	Human pancreatic c
34	145	4.9	40	21 AAB54409	Cat flea HWT allan
35	143	4.8	384	21 AAB29608	Klebsiella aerogen
36	142	4.8	567	19 AAW37774	Klebsiella aerogen
37	142	4.8	567	21 AAY81825	Helicobacter pylori
38	137.5	4.6	749	19 AAW80599	Part of protein wi
39	136.5	4.6	569	11 AAR04580	Helicobacter felis
40	136.5	4.6	589	16 AAR74337	H. felis structura
41	136.5	4.6	589	17 AAW06730	H. pylori urease B
42	136.5	4.6	569	17 AAW07194	H. pylori GHPO 124
43	136.5	4.6	569	19 AAW98511	Urease A and B sub
44	134.5	4.5	806	16 AAR67371	H. pylori urease u
45	133.5	4.5	569	16 AAR67378	

ALIGNMENTS

RESULT 1

AAW68489
ID AAW68489 standard; Protein: 553 AA.

XX
AC AAW68489;

XX
DT 08-DEC-1998 (first entry)

XX
DE Human partial ULIP-4 protein.

XX
Human; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis; neurodegenerative disorder; diagnosis.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT Misc-difference 56

FT /note= "encoded by TGA"

XX
PN FR2759701-A1.

XX
PD 21-AUG-1998.

XX
PF 19-FEB-1997; 97FR-0001961.

XX
PR 19-FEB-1997; 97FR-0001961.

XX
PA (INRM : INSERM INST NAT SANTE & RECH MEDICALE.

XX
PI Aquera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

XX
PI Quach TT, Sobel A;

XX
DR WPI: 1998-449610/39.

DR
N-PSDB; AAW60818.

```

XX PT Mouse and human ULIP poly:peptide(s) - useful in detection of
XX PT para-neoplastic neurological syndromes
XX PS Claim 1; Fig 12; 90pp; French.
XX CC This sequence represents a partial human Unc-33-like phospho-protein
XX CC (ULIP)-4. The coding sequence was isolated based on similarity to the
XX CC rat ULIP sequence. Proteins of the ULIP family or their corresponding
XX CC nucleic acids can be used in compositions for treating neurodegenerative
XX CC disorders and neoplasms, especially for para-neoplastic neurological
XX CC syndromes and/or for the early diagnosis of tumourigenesis.
XX SQ Sequence 553 AA;

Query Match 96.5%; Score 2882; DB 19; Length 553;
Best Local Similarity 99.8%; Pred. No. 1e-268;
Matches 552; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFQKKSPRITSDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 msfqqkspitrsdrlirggrivnddgsfyadvhvedglikqigenlivpggixtida 60

QY 61 HGLMVLPGGVVHTRLOMPVLGMPVLTADDPCQGTAAALAGGTTMLDHPDTPGVSLAAY 120
Db 61 hglmvlpggvvhtlrlompvlgmtaddpcqgtkaalaggttmildhvpdtpgvslaaay 120

QY 121 EQWRERADSAACCDYSLHVDITRWHESTKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
Db 121 eqwreradaaaccdyslhvdi trwhesikeelealvkegvnsflvfmaykdrccsdsg 180

QY 181 MYEIPSIIRDLGALAOVHAENGDIVEEQKRLLELGITGPEGHVLSPHEVEAEAVYRAV 240
Db 181 myeifsiirdlgalaqvhaengdi veeqkrllelgitgpeghvlshpeveaeavyrav 240

QY 241 TIAKQANCLYVTKVMSKGAADATAIAQAKRRGVVVFGEPTASLGTGSHYWSKNWAKAAA 300
Db 241 tiakqancplyvtkvmskgaadaiaqakrrgvvvfgeptaslgtdgshywsknwakaaa 300

QY 301 FVTSPVPNPDPPTADHLTCLLSSGDLQVTGSAHCTFTTAQKAVGKDNFALIPGEGTNGIEE 360
Db 301 fvtspvpnpdpptadhl tcllssgdlqvtgsahctfttaqkavgkdnfalipegtngiee 360

QY 361 RMSVWEKCVASCKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINWPKATKII 420
Db 361 rmsvwekcvasckmdenefvavtstnaakifnfyprkgrvavgsdadlvwnpkatkii 420

QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGREVPVPRKTFPDFVYK 480
Db 421 sakt hnlnevynifegvecr gapavvisqgrvaledgkmfvtpgagr evpvprktfpdfvyk 480

QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSQFS 540
Db 481 rikarnrlaeithgvprg lydgpvh evmvpa kpgsgaparascp gkisvppvrlnhqsqfs 540

QY 541 LSGSQADHDHARR 553
Db 541 lsgsqadhdharr 553

```

RESULT 2

AAW68488

ID AAW68488 standard; Protein; 572 AA.

XX AC AAW68488;

XX DT 08-DEC-1998 (first entry)

XX DE Mouse ULIP-4 protein.

XX KW Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;

XX NW neurodegenerative disorder; diagnosis.

XX OS Mus musculus.
 XX PN FR2759701-A1.
 XX PD 21-AUG-1998.
 XX PF 19-FEB-1997; 97FR-0001961.
 XX PR 19-FEB-1997; 97FR-0001961.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
 XX PI Quach TT, Sobel A;
 XX DR WPI; 1998-449610/39.
 XX DR N-PSDB; AAV60817.
 XX PT Mouse and human ULIP poly:peptide(s) - useful in detection of
 XX PT para-neoplastic neurological syndromes
 XX PS Claim 1; Fig 11; 90pp; French.
 XX CC This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-4.
 XX CC The coding sequence was isolated based on similarity to the rat ULIP
 XX CC sequence. Proteins of the ULIP family or their corresponding nucleic
 XX CC acids can be used in compositions for treating neurodegenerative
 XX CC disorders and neoplasms, especially for para-neoplastic neurological
 XX CC syndromes and/or for the early diagnosis of tumourigenesis.
 XX SQ Sequence 572 AA;

Query Match 93.9%; Score 2803; DB 19; Length 572;
 Best Local Similarity 92.7%; Pred. No. 4.5e-261;
 Matches 530; Conservative 22; Mismatches 20; Indels 0; Gaps 0;

QY 1 MSFQKKSPRITSDRLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
 Db 1 msfqqkspitrsdrlirggrivnddgsfyadvhvedglikqigenlivpggiktida 60

QY 61 HGLMVLPGGVVHTRLOMPVLGMPVLTADDPCQGTAAALAGGTTMLDHPDTPGVSLAAY 120
 Db 61 hglmvlpggvvhtlrlompvlgmtaddpcqgtkaalaggttmildhvpdtpgvslaaay 120

QY 121 EQWRERADSAACCDYSLHVDITRWHESTKEELEALVKEGVNSFLVFMAYKDRCCSDSQ 180
 Db 121 eqwreradaaaccdyslhvdi trwhesikeelealvkegvnsflvfmaykdrccsdsg 180

QY 181 MYEIPSIIRDLGALAOVHAENGDIVEEQKRLLELGITGPEGHVLSPHEVEAEAVYRAV 240
 Db 181 myeifsiirdlgalaqvhaengdi veeqkrllelgitgpeghvlshpeveaeavyrav 240

QY 241 TIAKQANCLYVTKVMSKGAADATAIAQAKRRGVVVFGEPTASLGTGSHYWSKNWAKAAA 300
 Db 241 tiakqancplyvtkvmskgaadamaqakrrgvvvfgeptaslgtdgshywsknwakaaa 300

QY 301 FVTSPVPNPDPPTADHLTCLLSSGDLQVTGSAHCTFTTAQKAVGKDNFALIPGEGTNGIEE 360
 Db 301 fvtspvpnpdpptadhl tcllssgdlqvtgsahctfttaqkavgkdnfalipegtngiee 360

QY 361 RMSVWEKCVASCKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINWPKATKII 420
 Db 361 rmsvwekcvasckmdenefvavtstnaakifnfyprkgrvavgsdadlvwnpkatkii 420

QY 421 SAKTHNLNVEYNIFEGVECRGAPAVVISQGRVALEDGKMFVTPGAGREVPVPRKTFPDFVYK 480
 Db 421 sakt hnlnevynifegvecr gapavvisqgrvaledgkmfvtpgagr evpvprktfpdfvyk 480

QY 481 RIKARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPGKISVPPVRLNHQSQFS 540
 Db 481 rikarnrlaeithgvprg lydgpvh evmvpa kpgsgaparascp gkisvppvrlnhqsqfs 540

QY 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 572
 DB 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 572

RESULT 3

AAW68486
 ID AAW68486 standard; Protein; 572 AA.

AC AAW68486;

DT 08-DEC-1998 (first entry)

XX Mouse ULIP-1 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;
 KW neurodegenerative disorder; diagnosis.

XX Mus musculus.

XX FR2759701-AL.

XX 21-AUG-1998.

PF 19-FEB-1997; 97FR-0001961.

PR 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

PI Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX N-PSDB; AAV60815.

XX Mouse and human ULIP poly:peptide(s) - useful in detection of

PT para-neoplastic neurological syndromes

XX Claim 1; Fig 9; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 572 AA;

Query Match 77.9%; Score 2325; DB 19; Length 572;

Best Local Similarity 75.3%; Pred. No. 5.2e-215;

Matches 430; Conservative 70; Mismatches 71; Indels 0; Gaps 0;

QY 1 MSFGKSTPRITSDRLIRGGRIVNDQSFVADHVEBGLIKOIGENLIVPGIHTIDA 60

DB 1 msyggknpitdrillkxgkivnddgsfyadiymedgllkqivlenlvp9gvtiea 60

QY 61 HGLMVLPGGVVDVHRLQMPVLGTMTPADDFCQGTKAALAGGTTMILDHVPFDTGVSLAAY 120

DB 61 hsrmlvpggidvhrfcmqdgmtsaddffggtkaalaggttmildhvvpepgtsllaaf 120

QY 121 EQWRERADSACCDYSLHVDITRHHESIKEEELALYKEGVNSFLVFMAYKRCQCSDSQ 180

DB 121 dqvrewadskscdcyslhvditewhkgigeamealvkhgvsflvymafkdrfqtidsq 180

QY 181 MYEFSTIRDLGALAOVHAENGDI VEEOKRLLELTGTGEGHVLGHPVEEAAYRAV 240

DB 181 iyevlsvridgaiaqvhagengdiaeaqqrilditgtpgehvlrpeveeaavnsri 240

QY 241 TIAQANCPLYVTVMKSGAADAIAQAKRGVVVFGEPTASLGTDGSHYWSKNWAKAAA 300

DB 241 tianqntcplvytkvmkpsaaeviaaqarkgtvygepitaslgtgshywsknwaaaa 300
 QY 301 FVTSPVPNPDPFTADHLTCLLSGDLQVTSAGHCTFTTAOKAVGKDNFALIPGCTNGIEE 360
 DB 301 fvtspplspdpftadhlclllsgdlqvtsahctftttaokavgkdnftlpegtngtee 360
 QY 361 RMSMVWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKII 420
 DB 361 rmsvwdkavvtgkmdenqfvavtstnaakfnlyprkgrisvgsdadlvwdpdsvktl 420
 QY 421 SAKTHNLNVEYNIFEVGECEGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPVYK 480
 DB 421 sakthnsaleynifegmecrgsplvvisqgkivledgtllhvtgsgryiprkpfdfvyk 480
 QY 481 RIKARNRLAEHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRLHQSSES 540
 DB 481 rikarsrlaelrgvprglydgpvcvavtpktvtpassaktspakqappvrvnlhqsffs 540
 QY 541 LSGSQADHIAARRTAQKIMAPPGRSNTSL 571
 DB 541 lsgsqadidnprttqrivappggranitsl 571

RESULT 4

AAW68487

ID AAW68487 standard; Protein; 572 AA.

AC AAW68487;

DT 08-DEC-1998 (first entry)

DE Mouse ULIP-3 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumorigenesis;
 KW neurodegenerative disorder; diagnosis.

XX Mus musculus.

XX FR2759701-AL.

XX 21-AUG-1998.

XX 19-FEB-1997; 97FR-0001961.

XX 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

XX Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX N-PSDB; AAV60816.

XX Mouse and human ULIP poly:peptide(s) - useful in detection of
 PT para-neoplastic neurological syndromes

XX Claim 1; Fig 10; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumorigenesis.

XX Sequence 572 AA;

Query Match

Best Local Similarity 72.1%; Score 2151; DB 19; Length 572;

Matches 394; Conservative 85; Mismatches 92; Indels 0; Gaps 0;

QY 1 MSFQGGKSIPTISDRLLIRGRIRVNDOSFYADVHVEDGLIKQIGENLIVPGGIHTIDA 60
Db 1 mshqgkksiphtsdrllirgrirvndosfyadvhvedglkqigenlivpggvtiea 60
QY 61 HGLMVLPGGVVHTRLQMPVLQMTFADDFCQCTKAALAGTMTLHDHVPDGTGVSLLAAAY 120
Db 61 ngmivpggldvntylqkpsqgmteaddffqgkaalaggttmiidhvvpepgsslltsf 120
QY 121 EOWRERADSAACCDYSLHVDITRWHSIEKEELEAVLKEKGVNSFLVFMAYKDRQCQSDSQ 180
Db 121 ekwheadtkscddyslhdvicswgydvreelevlqdkgvnsfvmaykldyqmsdsq 180
QY 181 MYEIFSIRDLGALACQVHAENGDIIVEEOKRLELIGITGPEGHVLSHPDEVEAEAVYRAV 240
Db 181 lyeatftklgavilvhaengdliaeqekrilemgitgpeghalsrpeeaeavfrai 240
QY 241 TIAGANCPPLYTKVMSKGAADIAQAQRGVVVVGEPIATSLDGTGSHYWSKNWAKAAA 300
Db 241 alagincpvyitkvmksaaadialarkkplvfgelplaaalslgtgdthyswnwaka 300
QY 301 FVTSPVNPDPPTDHLKCLSSGDLQVTGSAHCFTTAQKAVGKDNALIPETNGITEE 360
Db 301 fvtspplspdpptpylksliacgdlvtgsgnecpyslaqkavgkdnfllipegvngiee 360
QY 361 RMSWWEKCVASGKMDNEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINWPKATKII 420
Db 361 rmtvwdkavatkndengfvavtstnaakifnlyprkgrlavgsdadvlwdpdkmkti 420
QY 421 SAKTHNLNVEYNIFGVSCRCGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTFDFVYK 480
Db 421 takshksveynifegmechdsplvisgkivfedgnisvskmgtrfipkrpfpehlyq 480
QY 481 RIKARNLAEIHGVPRGLYDGVHEVMVPAKPGSGAPARACPCGKISVPPVNLHQSFGS 540
Db 481 rvlrskvfglshvrgmgydpvvevpatpkhaapapsaespskhqppplrnlhqsfnfs 540
QY 541 LSGSQADHIAHRAQKIMAPGGRSNTSL 571
Db 541 lsgsqadhnprtrghrivappgrsntsl 571

RESULT 5
AAG46568
ID AAG46568 standard; Protein; 683 AA.
XX AC AAG46568;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58600.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 31.7%; Score 945; DB 21; Length 588;
Best Local Similarity 39.4%; Pred. No. 5,1e-82;


```
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145080.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
```

```
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159384.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 30.2%; Score 901; DB 21; Length 576;

Best Local Similarity 38.9%; Pred. No. 8.6e-78;

Matches 195; Conservative 86; Mismatches 178; Indels 42; Gaps 10;

```
QY 78 MEVLGWTADDQCCTKAALAGGTTMILDRHVPDPGVSLAAYEQWRERADSAACCDYSL 137
DB 1 mefmtetiddffsgaaalaggttmhldfvpvng-nlvagfeay-enkrescmdygf 58
QY 138 HVDITRWHESKEELEALVKGVNSFLVFMAYKDRCCSQSOMYEIFSIINDLGALAOV 197
DB 59 hmaatkdegvsrdmelmvkeginsfkflaykgslnvtdldllleglkrcksigalamv 118
QY 198 HAENGDIVEEQKRLLLELGITGPEGHVLSHPPEVEAEAVYRAVTTAKANCPLYVTKWS 257
DB 119 haengdavfegqkrmiekgitgpeghalsrppvlegatarairlarfintplyvvhms 178
QY 258 KCAADAIQAOKRRGVVFEGETASIGTDSHYWSKNNAKAAAFVTSPPVNDPTTADH- 316
DB 179 vdamdeiakarksgkvigepvvsghliddhwlwdpfdfiaskyvmppirp-vgbg 234
QY 317 --LTCLLSSGDLQVTSAGHCTFTTAAKAVGKDNFALIPETNGIEMRSMWKECVASGK 374
DB 235 kalqdalstgilvgttdhctfnstqkalgldfrpbgvngleermhliwdtmvesgg 294
QY 375 MDNEFVAVTSTNAKIFNFPKRGVAVGSDADLVINPNKATKIISAKTHNLNVEYNIF 434
DB 295 lsaatyvritstecarifniyprkgailagsdadliilpnssyelskskshrsdtnvy 354
QY 435 EGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTFPDPFVVKRIK----- 483
DB 355 egrrgkgvevtiaggrivweneelkvprsgkytemppf-sylfdeeekeetceklslin 413
QY 484 ---ARNRLAEINGVPRGLYDGPVHEVMVPAKPGCSGAPARASC-----PGKISVPP- 530
DB 414 itvsnnqafeleaqaqknffhggdnk--spankgrkrrreamtittslrksntrlppevnr 471
QY 531 ---VRNLHOSGFSLSGSQADD 548
DB 472 vlyvrnl---pfnitseeemdyd 489
```

RESULT 9

AAR29715 AAR29715 standard; Protein; 471 AA.

XX AAR29715;

XX AAR29715;

DT 05-MAY-1993 (first entry)

XX Heat stable hydantoinase protein.

XX Hydantoinease; E.coli; heat; stable; thermal.
 XX Bacillus stearothermophilus.
 OS JP04325093-A.
 PN 13-NOV-1992.
 PD 23-APR-1991; 91JP-0117802.
 PF 23-APR-1991; 91JP-0117802.
 PR (NIPS) NIPPON SODA CO.
 XX WPI; 1992-428828/52.
 DR N-PSDB; AAQ31987.
 XX Heat-stable hydantoinease (I) gene - has specified base and amino acid sequence and is prepd. from E. coli
 PT Claim 2; Page 5-7; 8pp; Japanese.
 PS This sequence represents a heat-stable hydantoinease protein. The DNA encoding this peptide can be used to transform E.coli such that the protein is produced stably by the microbe.
 CC Sequence 471 AA;
 SQ

Query Match 28.08; Score 837; DB 13; Length 471;
 Best Local Similarity 40.08; Pred. No. 9e-72;
 Matches 183; Conservative 79; Mismatches 192; Indels 4; Gaps 4;

QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGGVVDVHRLQ 77
 Db 4 likngtvtatdiyeaddllldqgiavgrnl-desgaevidatgcyvfpvggldphthld 62
 QY 78 MPVLGMPADDFCOGTAALAGGTTMILDHVPDGTGVSLLAAYEQWRERADSAACCDYSL 137
 Db 63 mpfggtvtkddsfesgtiaaafgggtttidfcitnkgelkkaletwhnkatgkavidygf 122
 QY 138 HVDITRWHESIKELEALVKEKGVNSFLVFMAYKRCQCSQSMYEIFSIIRDLGALAQV 197
 Db 123 hlmseitddvleelpkvieegitsfkvmaykdvfgaddgtlyrtlvaaakelgalvmv 182
 QY 198 HAENGDIVVEEQKRLLEIGITGPEGHVLSHPPEEVEAEAVYRAVVTIAKOANCPLYVTKVMS 257
 Db 183 haengdivdyitkkaledghdtplyhaltrppeleageatgracqltelagsqlyvvhsc 242
 QY 258 KGADATAQAQRKRGVVVEGPEITASLGTDSGHYKSNWAKAAAFVTPPVNPDPTTADHL 317
 Db 243 aqavekiaearnkglvnmwgetcpqylvldgsylekpnf-egakyvwpplr-ekwhqevl 300
 QY 318 TCLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPGTNGIEERMSVMWEKCVASGKMD 376
 Db 301 wnalngqltldgsdqcsfdkqgelgrdftkipnggpliedrvsilfsegvkkgrit 360
 QY 377 ENFVAVTSTNAAKIFNFPYPRKGVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFE 436
 Db 361 lngfdivstriaklflgfpkgtiavgadadlvifdptervisaeathmavdnpfeg 420
 QY 437 VECCGAPAVISQGRVALEDGKMFVETPGAGRFVPRKTF 474
 Db 421 mkvtgepvsvlcrgefvrdrkqfvgkpgygyvkraky 458

RESULT 10
 AAR69874
 ID AAR69874 standard; Protein; 460 AA.
 XX
 AC AAR69874;
 XX

DT 27-OCT-1995 (first entry)
 XX B.thermoglucoasius D-hydantoinease with mutated C-terminal region.
 DE D-hydantoinease; dihydropyrimidinase; mutein; variant; mutation;
 KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;
 KW production; racemisation.
 XX Bacillus thermoglucoasius.
 OS DE4328829-A.
 PN 02-MAR-1995.
 PD 27-AUG-1993; 93DE-4328829.
 PF 27-AUG-1993; 93DE-4328829.
 PR (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA Burtscher H, Lang G, Popp F;
 PI WPI; 1995-099331/14.
 DR N-PSDB; AAQ80565.
 XX New highly active and stable D-hydantoinease enzyme - isolated from Bacillus thermoglucoasius, useful in the prodn. of D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin
 PT Claim 1; Page 5-6; 8pp; German.
 PS The wild-type D-hydantoinease gene was isolated from Bacillus thermoglucoasius using the amplification primers Hyd1 and Hyd2 (see AAQ80566 and AAQ80567). A single-base deletion in the HindIII site (from AACCTT to AACCT) was then introduced to give a coding region (AAQ80565) that encoded a D-hydantoinease (AAR69874) that was shorter than the wild-type protein with a different C-terminal sequence (see Features Table). The mutant enzyme was found to have improved heat stability and higher activity compared to known hydantoineases.
 CC Sequence 460 AA;
 SQ

Query Match 27.7%; Score 826; DB 16; Length 460;
 Best Local Similarity 39.7%; Pred. No. 1e-70;
 Matches 182; Conservative 78; Mismatches 194; Indels 4; Gaps 4;

QY 18 LIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGIHTIDAHGLMVLPGGVVDVHRLQ 77
 Db 4 likngtvtatdiyeaddllldqgiavgrnl-desgaevidatgcyvfpvggldphthld 62
 QY 78 MPVLGMPADDFCOGTAALAGGTTMILDHVPDGTGVSLLAAYEQWRERADSAACCDYSL 137
 Db 63 mpfggtvtkddsfesgtiaaafgggtttidfcitnkgelkkaletwhnkatgkavidygf 122
 QY 138 HVDITRWHESIKELEALVKEKGVNSFLVFMAYKRCQCSQSMYEIFSIIRDLGALAQV 197
 Db 123 hlmseitddvleelpkvieegitsfkvmaykdvfgaddgtlyrtlvaaakelgalvmv 182
 QY 198 HAENGDIVVEEQKRLLEIGITGPEGHVLSHPPEEVEAEAVYRAVVTIAKOANCPLYVTKVMS 257
 Db 183 haengdivdyitkkaledghdtplyhaltrppeleageatgracqltelagsqlyvvhsc 242
 QY 258 KGADATAQAQRKRGVVVEGPEITASLGTDSGHYKSNWAKAAAFVTPPVNPDPTTADHL 317
 Db 243 aqavekiaearnkglvnmwgetcpqylvldgsylekpnf-egakyvwpplr-ekwhqevl 300
 QY 318 TCLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPGTNGIEERMSVMWEKCVASGKMD 376
 Db 301 wnalngqltldgsdqcsfdkqgelgrdftkipnggpliedrvsilfsegvkkgrit 360
 QY 377 ENFVAVTSTNAAKIFNFPYPRKGVAVGSDADLVINWPKATKIISAKTHNLNVEYNIFE 436
 Db 421 mkvtgepvsvlcrgefvrdrkqfvgkpgygyvkraky 458

Db 361 lncfvdinmstriaklfglfrkgtiavgsdadlviifddiervisaethmavdynafeg 420
QY 437 VECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 421 mkvtgepvsvlrcgefvvrdkqfvgkpgygyllkagcf 458
RESULT 11
AAW02837
ID AAW02837 standard; Protein: 457 AA.
XX
AC AAW02837;
XX
DT 13-MAR-1996 (first entry)
XX
DE Agrobacterium radiobacter D-hydantoinase.
XX
KW D-hydantoinase; D-N-carbamylase; enzyme; stereospecific reaction;
KW D-amino acid.
XX
OS Agrobacterium radiobacter.
XX
PN EP677585-A1.
XX
PD 18-OCT-1995.
XX
PF 24-MAR-1995; 95EP-0104393.
XX
PR 15-APR-1994; 94IT-OMI0726.
XX
PA (ENTIE) ENRICERCH SPA.
XX
PI Frascotti G, Galli G, Grandi G, Grifantini R;
XX
DR WPI; 1995-352764/46.
DR N-PSDB; AAT01498.
XX
PT Prodn. of D-alpha amino acids from racemic 5-substd. hydantoin cpds.
PT - using microorganisms contg. hydantoinase and carbamylase genes.
XX
PS Disclosure; Fig.5A-5C; 44pp; English.
XX
CC A. radiobacter is the donor microorganism for genes encoding
CC D-hydantoinase and D-N-carbamylase which are expressed in
CC Escherichia coli or Bacillus subtilis using plasmid pSM651. The
CC resulting recombinant E. coli may be used to catalyze the
CC stereospecific preparation of D-amino acids from racemic
CC 5-substituted hydantoin compounds.
XX
SQ Sequence 457 AA;
Query Match 23.2%; Score 691.5; DB 16; Length 457;
Best Local Similarity 35.3%; Pred. NO. 9e-58;
Matches 165; Conservative 79; Mismatches 200; Indels 23; Gaps 9;
QY 17 LLJRGGRIVNDQSFADYHVEGLIKQIGENLIVPGGIIHDAHGLMVLPGGVVDVHTL 76
Db 3 lllkngtvtadvgsdpdgkdgkkaqig-9tfgpag-rtidasgryvfp9g9dvhtv 60
QY 77 QMPVLGMPADDFCQGTAKAALAGTTMTILDHVPDP7GVSLAAAYEQWRERADSAACCDYS 136
Db 61 etvsfnqtsadtatvaacggtttvdvfcqdgghsireavakwdgmagkksaidyg 120
QY 137 LHVDITRWHSIEEELAEVKEGVNSFLVPMAYKRCQCSQSDSQMYEIFSIIRDGLALAQ 196
Db 121 yhiivlptdsvleeelevl-pdligitsfkvmayrgmmidvtilrtldkaaktgslvm 179
QY 197 VHAENGDIVEEOKRLELGITGPEGHVLSHPEVEAEAYVRAVTIAKQANCPLYVTYKVM 256
Db 180 vhaengdaadyldrkfvadgktaapiyhalrpprveaeataralalaevnapiyihlt 239
QY 257 SKGAADAIAQAKRRGVVVFGEPIITASIGTDGSHYWSKNWA-----KAAAFVTPSPVNDP 311

Db 240 ceasfdelmtrakargvhalaetctqyl-----yltkddlerpdefagkyvftpp-----p 289
QY 312 TTADHLTCL---LSSGDLQVTGSAHCTFT-TAQKAVGKDNFALIPETNGIEERMSMWWE 367
Db 290 rtkkdqellwnalnrgvletvssdhcswlfeghkdrgrndfraipngapgvveerlmmyq 349
QY 368 KCVASGKMDENEFVAVTSTNAAKIFNFYPRKGRVAVGSDADLVINWPKATKIISAKTHNL 427
Db 350 g-vnegrisltqfvelvatrpakvfgmfpekgvtavgsdddivlwdpeaemviegsmah 408
QY 428 NVEYNIIEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRFVPRKTF 474
Db 409 amdyssyeghkikgvpktvllrgkvivdegtyvgagtdgqfkrky 455
RESULT 12
AAW03023
ID AAW03023 standard; Protein: 457 AA.
XX
AC AAW03023;
XX
DT 31-JAN-1997 (first entry)
XX
DE Bacillus sp. hydantoinase.
XX
KW Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
KW intermediate; drug synthesis; penicillin; cephalosporin;
KW antibiotic.
XX
OS Bacillus sp.
XX
PN WO9620275-A1.
XX
PD 04-JUL-1996.
XX
PF 26-DEC-1995; 95WO-JP02688.
XX
PR 28-DEC-1994; 94JP-0326865.
XX
PA (KANF) KANEGAFUCHI KAGAKU KOGYO KK.
XX
PI Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
PI Yamada Y;
XX
DR WPI; 1996-321848/32.
DR N-PSDB; AAT31256.
XX
PT Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
PT a recombinant hydantoinase derived from a strain of Pseudomonas,
PT Agrobacterium or Bacillus
XX
PS Claim 15; Page 25-29; 54pp; English.
XX
CC D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
CC hydantoin by treatment with a hydantoinase expressed by a
CC transformant microorganism carrying a vector containing DNA coding
CC for the hydantoinase and derived from Bacillus sp. KNK245,
CC Agrobacterium sp. KNK712 or Pseudomonas sp. KNK003A. The D-N-
CC carbamoyl-alpha-amino acid can be used for the production of
CC optically active alpha amino acids (especially D-phenylglycine and D-
CC p-hydroxyphenylglycine) as intermediates for drug synthesis,
CC especially for the production of semi-synthetic penicillin and
CC cephalosporin antibiotics.
XX
SQ Sequence 457 AA;
Query Match 22.9%; Score 684.5; DB 17; Length 457;
Best Local Similarity 34.3%; Pred. NO. 4.2e-57;
Matches 161; Conservative 83; Mismatches 197; Indels 29; Gaps 9;
QY 17 LLJRGGRIVNDQSFADYHVEGLIKQIGENLIVPGGI-----HTIDAGLAVLPGGVVDV 72

Db 3 iikngtvtadgisradigikdgkitqg-----gaigpaertidaagryvfpqgidv 56
 QY 73 HTRLOPVLGMPADDFCOGQTKAALAGGTMILDHVFPDTGVSLLAAYEQWRERADSAAC 132
 Db 57 hthvetsvntqsadtfatavaacgggtttivdfcqqdgrghslaepvkvkwdmaggksa 116
 QY 133 CDYSLHVDITRWHESKEELEALVKEKGVNSFLVFMAYKDRQCQSDSQMYEIFSIIIRDLG 192
 Db 117 idygyhiivldptdsvieeevl-pdlgitsfkvmayrgmmiddtllktldkavktg 175
 QY 193 ALAQVHAENGDIIVEEOKRLLELGTGPGHVLSPHEVEAEAVRAVTVIAQANCPPLV 252
 Db 176 slvmvhaengdaadyldrkfvaegktapiyhalssrprveaeataralaleivnapiyi 235
 QY 253 TKVMSKGAADAIQAOKRRGVVVFGEPTITASLGTGDSHYWSKNWA-----KAAAFVTSPPV 307
 Db 236 vhtvteesleevmraksgvralactcthyt-----ytkedlerpdegakvftppa 289
 QY 308 NPDPTTADH--LTCLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPETNGIERMSM 364
 Db 290 r---akkdhvlnalrngvftvssdhcswlfkghkdrgrndfraipngapgveerlmm 346
 QY 365 VWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKT 424
 Db 347 vyqg-vnegrialtqfvelvatrpakvfgmpqkctiavgsdadivldpeaemvieqta 405
 QY 425 HNLNVEYNIFEGVEGCRGAPAVVISQGRVALEDGKMFVTPGAGRFPVPRKTF 474
 Db 406 mnamdyssyeghkvkvpktvllrgkvivdegsvygeptdgkfkkrky 455

RESULT 13

AAW03024
 ID AAW03024 standard; Protein; 457 AA.

XX AC AAW03024;

XX DT 31-JAN-1997 (first entry)

XX DE Agrobacterium hydantoinase.

XX KW Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
 KW intermediate; drug synthesis; penicillin; cephalosporin;
 XX antibiotic.

XX OS Agrobacterium sp.

XX FN WO9620275-A1.

XX PD 04-JUL-1996.

XX PF 26-DEC-1995; 95WO-JP02688.

XX PR 28-DEC-1994; 94JP-0326865.

XX PA (KANF) KANEGAFUCHI KAGAKU KOGYO KK.

XX PI Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
 PI Yamada Y;

XX XX WPI; 1996-321848/32.

XX DR N-PSDB; AAT31257.

XX PT Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
 PT a recombinant hydantoinase derived from a strain of Pseudomonas,
 PT Agrobacterium or Bacillus

XX PS Claim 15; Page 30-34; 54pp; English.

XX CC D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted

XX CC hydantoin by treatment with a hydantoinase expressed by a

XX CC transformant microorganism carrying a vector containing DNA coding

QC

CC for the hydantoinase and derived from Bacillus sp. KNK245.
 CC Agrobacterium sp. KNK712 or Pseudomonas sp. KNK003A. The D-N-
 CC carbamoyl-alpha-amino acid can be used for the production of
 CC optically active alpha amino acids (especially D-phenylglycine and D-
 CC p-hydroxyphenylglycine) as intermediates for drug synthesis,
 CC especially for the production of semi-synthetic penicillin and
 CC cephalosporin antibiotics.

XX Sequence 457 AA;

Query Match 22.9%; Score 684.5; DB 17; Length 457;
 Best Local Similarity 34.3%; Pred. No. 4-2e-57;
 Matches 161; Conservative 83; Mismatches 197; Indels 29; Gaps 9;

QY 17 LLIRGRIVNDQSFYADVHVEDGLIKQIGENLIVPGGI-----HTIDAHLMLVPGVDV 72
 Db 3 iikngtvtadgisradigikdgkitqg-----gaigpaertidaagryvfpqgidv 56
 QY 73 HTRLOPVLGMPADDFCOGQTKAALAGGTMILDHVFPDTGVSLLAAYEQWRERADSAAC 132
 Db 57 hthvetsvntqsadtfatavaacgggtttivdfcqqdgrghslaepvkvkwdmaggksa 116
 QY 133 CDYSLHVDITRWHESKEELEALVKEKGVNSFLVFMAYKDRQCQSDSQMYEIFSIIIRDLG 192
 Db 117 idygyhiivldptdsvieeevl-pdlgitsfkvmayrgmmiddtllktldkavktg 175
 QY 193 ALAQVHAENGDIIVEEOKRLLELGTGPGHVLSPHEVEAEAVRAVTVIAQANCPPLV 252
 Db 176 slvmvhaengdaadyldrkfvaegktapiyhalssrprveaeataralaleivnapiyi 235
 QY 253 TKVMSKGAADAIQAOKRRGVVVFGEPTITASLGTGDSHYWSKNWA-----KAAAFVTSPPV 307
 Db 236 vhtvteesleevmraksgvralactcthyt-----ytkedlerpdegakvftppa 289
 QY 308 NPDPTTADH--LTCLLSSGDLQVTSAGHCTFT-TAQKAVGKDNFALIPETNGIERMSM 364
 Db 290 r---akkdhvlnalrngvftvssdhcswlfkghkdrgrndfraipngapgveerlmm 346
 QY 365 VWEKCVASGKMDENEFVAVTSTNAKIFNFYPRKGRVAVGSDADLVINPKATKIISAKT 424
 Db 347 vyqg-vnegrialtqfvelvatrpakvfgmpqkctiavgsdadivldpeaemvieqta 405
 QY 425 HNLNVEYNIFEGVEGCRGAPAVVISQGRVALEDGKMFVTPGAGRFPVPRKTF 474
 Db 406 mnamdyssyeghkvkvpktvllrgkvivdegsvygeptdgkfkkrky 455

RESULT 14

AAAB58948

ID AAB58948 standard; Protein; 167 AA.

XX AC AAB58948;

XX DT 27-MAR-2001 (first entry)

XX DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 656.

XX KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neoplastic; neuroprotective; antiviral; antitumor; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX OS Homo sapiens.

XX PN WO200055173-A1.

XX XX 21-SEP-2000.

XX PD

PF 08-MAR-2000; 2000WO-US05881.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX PI
XX WFI; 2000-611515/58.
DR N-PSDB; AAF21851.
XX
XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
XX Claim 11; Page 1102; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 167 AA;
SQ

Query Match 20.1%; Score 600; DB 21; Length 167;
Best Local Similarity 68.1%; Pred. No. 1.1e-49;
Matches 113; Conservative 22; Mismatches 31; Indels 0; Gaps 0;
QY 406 DADLVINPKATKIISAKTHNLNVEYIFEGVECRGAPAVVISQGRVALEDGRKFVTPGA 465
Db 1 dadiwvdpdsvktisakthnslsleyifegmeorgsplsvisgqkivldgtlhtvtxs 60
QY 466 GRFVPRKTFDFVYKRIKARNLAEIHGVPRLYDGPVHVMVPKPGSGAPARASCPGK 525
Db 61 gryiprkpfdfykyrikarslaelrgvprglydgvcevsvtptvtpassaktspak 120
QY 526 ISVPPVRLNHQSFGSLSGSQADPHIARRTAOKIMAPPGRSNTLSL 571
Db 121 qqappvnlhqsgfslsgaiddnprtttqtrivappggranitsl 166

RESULT 15
AAG23637
ID AAG23637 standard; Protein; 281 AA.
XX
XX AAG23637;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 27019.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.

XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127482.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 30, 2001, 11:40:36 ; Search time 21.48 seconds
(without alignments)
1614.382 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFGKKSIPRTSDLLR.....RTAKTMAPGGRSNITSLS 572

Scoring table: OLIGO 60.0 , Gapext 60.0

Searched: 412676 seqs, 60623988 residues

Word size : 0

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : A_Geneseq_0601.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	497	86.9	553	19 AAW68489	Human partial ULIP
2	50	8.7	572	19 AAW68488	Mouse ULIP-4 prote
3	31	5.4	572	19 AAW68486	Mouse ULIP-1 prote
4	18	3.1	572	19 AAW68487	Mouse ULIP-3 prote
5	15	2.6	167	21 AAB58948	Breast and ovarian
6	13	2.3	40	21 AAB54409	Human pancreatic c
7	12	2.1	27	19 AAW69259	DMS protein fragme
8	11	1.9	20	19 AAW69260	DMS protein fragme
9	11	1.9	20	19 AAW69262	DMS protein fragme
10	10	1.7	460	16 AAR69874	B.thermoglucodasiu
11	10	1.7	485	17 AAW03025	Pseudomonas hydant

12	10	1.7	576	21 AAG46570	Arabidopsis thalia
13	10	1.7	588	21 AAG46569	Arabidopsis thalia
14	10	1.7	588	21 AAG46568	Arabidopsis thalia
15	9	1.6	283	21 AAG23636	Arabidopsis thalia
16	9	1.6	284	21 AAG23635	Arabidopsis thalia
17	8	1.4	169	21 AAB43021	Agrobacterium radi
18	8	1.4	457	16 AAR82837	Streptomyces chole
19	8	1.4	526	15 AAR60325	Cholesterol oxidas
20	8	1.4	546	17 AAW08932	Cholesterol oxidas
21	8	1.4	546	17 AAW08933	Cholesterol oxidas
22	8	1.4	546	17 AAW08934	Cholesterol oxidas
23	8	1.4	546	17 AAW08935	Cholesterol oxidas
24	8	1.4	546	17 AAW08936	Cholesterol oxidas
25	8	1.4	546	17 AAW08930	Wild type choleste
26	3	1.4	546	17 AAW08931	Cholesterol oxidas
27	3	1.4	837	20 AAW75425	Human aggreacan deg
28	3	1.4	837	21 AAY99429	Human PRO1563 (UNQ
29	8	1.4	837	22 AAB66178	Protein of the inv
30	7	1.2	15	18 AAW10617	L-galactono-1,4-la
31	7	1.2	55	19 AAW69340	Secreted protein o
32	7	1.2	90	21 AAG33074	Arabidopsis thalia
33	7	1.2	90	21 AAG48928	Arabidopsis thalia
34	7	1.2	91	21 AAG33073	Arabidopsis thalia
35	7	1.2	91	21 AAG48927	Arabidopsis thalia
36	7	1.2	93	22 AAB80092	Corynebacterium gl
37	7	1.2	110	19 AAW63722	C. histolyticum CH
38	7	1.2	121	21 AAG24039	Arabidopsis thalia
39	7	1.2	124	18 AAW20410	H. pylori cytoplas
40	7	1.2	124	18 AAW24654	H. pylori cytoplas
41	7	1.2	125	15 AAR62920	Human cytomagalovi
42	7	1.2	125	15 AAR62918	Human cytomagalovi
43	7	1.2	126	14 AAR36700	Flagellin (amino a
44	7	1.2	134	21 AAG24038	Arabidopsis thalia
45	7	1.2	139	21 AAY43865	Heavy chain (VH) 9

ALIGNMENTS

RESULT 1
AAW68489
ID AAW68489 standard; Protein: 553 AA.

- XX AAW68489;
- XX 08-DEC-1998 (first entry)
- XX Human partial ULIP-4 protein.
- XX DE Human; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
- XX KW neurodegenerative disorder; diagnosis.
- XX OS Homo sapiens.
- XX Key Location/Qualifiers
- XX FT Misc-difference 56 /note- "encoded by TGA"
- XX FT
- XX FR2759701-A1.
- XX PN
- XX PD 21-AUG-1998.
- XX PF 19-FEB-1997; 97FR-0001961.
- XX PR 19-FEB-1997; 97FR-0001961.
- XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
- XX PI Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;
- XX PI Quach TT, Sobel A;
- XX DR WPI; 1998-449610/39.
- XX DR N-PSDB; AAW60818.

Claim 1; Fig 9; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-1.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumourigenesis.

XX Sequence 572 AA;

Query Match 5.4%; Score 31; DB 19; Length 572;

Best Local Similarity 100.0%; Pred. No. 1.2e-21;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPTASLGTDGSHYWSKNWAKAAAFVTSPP 306

Db 276 gepitaslgtgdshywsknwakaafvtspp 306

RESULT 4

AAW68487
 ID AAW68487 standard; Protein; 572 AA.

XX AC

XX AC

XX 08-DEC-1998 (first entry)

XX DE

XX Mouse ULIP-3 protein.

XX Mouse; Unc-33-like phospho-protein; ULIP; rat; neoplasm; tumourigenesis;
 KW neurodegenerative disorder; diagnosis.

XX OS

XX Mus musculus.

XX FR2759701-A1.

XX 21-AUG-1998.

XX 19-FEB-1997; 97FR-0001961.

XX 19-FEB-1997; 97FR-0001961.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Aguera M, Belin MF, Byk T, Honnorat J, Kolattukudy P;

XX Quach TT, Sobel A;

XX WPI; 1998-449610/39.

XX N-PSDB; AAV60816..

XX Mouse and human ULIP poly:peptide(s) - useful in detection of

XX para-neoplastic neurological syndromes

XX Claim 1; Fig 10; 90pp; French.

XX This sequence represents the mouse Unc-33-like phospho-protein (ULIP)-3.
 CC The coding sequence was isolated based on similarity to the rat ULIP
 CC sequence. Proteins of the ULIP family or their corresponding nucleic
 CC acids can be used in compositions for treating neurodegenerative
 CC disorders and neoplasms, especially for para-neoplastic neurological
 CC syndromes and/or for the early diagnosis of tumourigenesis.

XX Sequence 572 AA;

Query Match 3.1%; Score 18; DB 19; Length 572;

Best Local Similarity 100.0%; Pred. No. 6.2e-09;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306

|||||

Db 289 hywsknwakaafvtspp 306

RESULT 5

AAW58948

ID AAW58948 standard; Protein; 167 AA.

XX AAW58948;

XX 27-MAR-2001 (first entry)

XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 656.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.

XX WO20005173-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

XX N-PSDB; AAF21851.

XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -

PS Claim 11; Page 1102; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAW58711 - AAW59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAW59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 167 AA;

Query Match 2.6%; Score 15; DB 21; Length 167;

Best Local Similarity 100.0%; Pred. No. 1.8e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 PPVRNLHQSGFSLG 543
 Db 124 ppvrnlhqsgfslg 138

RESULT 6

AAW54409
 ID AAB54409 standard; Protein; 40 AA.

XX AC AAB54409;

XX DT 09-MAR-2001 (first entry)

XX XX Human pancreatic cancer antigen protein sequence SEQ ID NO:861.

XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

XX OS Homo sapiens.

XX XX WO200055320-A1.

XX PN 21-SEP-2000.

XX PD 08-MAR-2000; 2000WO-US03989.

XX PF 12-MAR-1999; 99US-0124270.

XX PR (HUMA-) HUMAN GENOME SCI INC.

XX PA Rosen CA, Ruben SM;

XX PI WPI; 2000-579444/54.

XX DR N-PSDB; AAC99174.

XX XX New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -

XX PS Claim 11; Page 1320; 1379pp; English.

XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 40 AA;

Query Match

2.3%; Score 13; DB 21; Length 40;

Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 QGTRKAALAGGTTM 103

Db 26 qgtkaalaggttm 38

RESULT 7

AAW69259
 ID AAW69259 standard; peptide; 27 AA.

XX AC AAW69259;

XX DT 29-OCT-1998 (first entry)

XX XX DMS protein fragment #1.

XX KW DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
 KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
 KW Alzheimer's disease; senile amyloid plaque.

XX OS Homo sapiens.

XX XX WO9834643-A1.

XX PN 13-AUG-1998.

XX PD 06-FEB-1998; 98WO-CA00065.

XX PF 03-FEB-1998; 98US-0017689.

XX PR 07-FEB-1997; 97US-0038694.

XX XX (NYMO-) NYMOX CORP.

XX PI Averbach P;

XX DR WPI; 1998-446954/38.

XX XX Medicaments preventing the formation of cerebral amyloid of the
 PT brain - by inhibiting synthesis, growth and disruption of dense
 PT microspheres, useful for preventing or treating Alzheimer's disease

XX PS Claim 3; Page 23; 43pp; English.

XX CC This sequence represents a fragment of the dense microsphere (DMS)
 CC protein. This sequence is recognised by an antibody used in the
 CC medicament used in the method of the invention. The method is for
 CC preventing the formation of cerebral amyloid in the mammalian brain, and
 CC comprises administering a medicament that prevents or inhibits the
 CC synthesis, growth and/or disruption of DMS. The method is used for the
 CC prevention, treatment or the prophylaxis of Alzheimer's disease by
 CC impeding the formation of senile amyloid plaques.

XX SQ Sequence 27 AA;

Query Match

2.1%; Score 12; DB 19; Length 27;
 Best Local Similarity 100.0%; Pred. No. 0.00032;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 LGITGPEGHVLIS 226

Db 4 lgitgpeghvlis 15

RESULT 8

AAW69260
 ID AAW69260 standard; peptide; 18 AA.

XX AC AAW69260;

XX XX 29-OCT-1998 (first entry)

XX DMS protein fragment #2.
 DE DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
 KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
 KW Alzheimer's disease; senile amyloid plaque.
 XX Homo sapiens.
 OS WO9834643-A1.
 PN 13-AUG-1998.
 PD 06-FEB-1998; 98WO-CA00065.
 PF 03-FEB-1998; 98US-0017689.
 XX 07-FEB-1997; 97US-0038694.
 PR (NYMO-) NYMOX CORP.
 PA Averbach P;
 XX WPI; 1998-446954/38.
 XX Medicaments preventing the formation of cerebral amyloid of the
 PT brain - by inhibiting synthesis, growth and disruption of dense
 PT microspheres, useful for preventing or treating Alzheimer's disease
 XX Claim 3; Page 23; 43pp; English.
 XX This sequence represents a fragment of the dense microsphere (DMS)
 CC protein. This sequence is recognised by an antibody used in the
 CC medicament used in the method of the invention. The method is for
 CC preventing the formation of cerebral amyloid in the mammalian brain, and
 CC comprises administering a medicament that prevents or inhibits the
 CC synthesis, growth and/or disruption of DMS. The method is used for the
 CC prevention, treatment or the prophylaxis of Alzheimer's disease by
 CC impeding the formation of senile amyloid plaques.
 XX Sequence 18 AA;
 SQ

Query Match 1.9%; Score 11; DB 19; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 AVGSDADLVIV 412
 DB 2 avgdsadlviv 12
 |||||

RESULT 9
 AAW69262
 ID AAW69262 standard; peptide; 20 AA.
 XX AC
 XX AAW69262;
 XX 29-OCT-1998 (first entry)
 DE DMS protein fragment #4.
 XX DMS protein fragment; dense microsphere; DMS; cerebral amyloid formation;
 KW antibody recognition sequence; mammalian brain; inhibitor; therapy;
 KW Alzheimer's disease; senile amyloid plaque.
 XX Homo sapiens.
 OS WO9834643-A1.
 PN 13-AUG-1998.
 PD 06-FEB-1998; 98WO-CA00065.
 PF

PR 03-FEB-1998; 98US-0017689.
 PR 07-FEB-1997; 97US-0038694.
 XX (NYMO-) NYMOX CORP.
 PA Averbach P;
 XX WPI; 1998-446954/38.
 XX Medicaments preventing the formation of cerebral amyloid of the
 PT brain - by inhibiting synthesis, growth and disruption of dense
 PT microspheres, useful for preventing or treating Alzheimer's disease
 XX Claim 3; Page 23; 43pp; English.
 XX This sequence represents a fragment of the dense microsphere (DMS)
 CC protein. This sequence is recognised by an antibody used in the
 CC medicament used in the method of the invention. The method is for
 CC preventing the formation of cerebral amyloid in the mammalian brain, and
 CC comprises administering a medicament that prevents or inhibits the
 CC synthesis, growth and/or disruption of DMS. The method is used for the
 CC prevention, treatment or the prophylaxis of Alzheimer's disease by
 CC impeding the formation of senile amyloid plaques.
 XX Sequence 20 AA;
 SQ

Query Match 1.9%; Score 11; DB 19; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.0024;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 IVNDQSFYAD 34
 DB 1 ivndqsfyad 11
 |||||

RESULT 10
 AAR69874
 ID AAR69874 standard; Protein; 460 AA.
 XX AC
 XX AAR69874;
 XX 27-OCT-1995 (first entry)
 XX B.thermoglucodasius D-hydantoinase with mutated C-terminal region.
 DE D-hydantoinase; dihydropyrimidinase; mutein; variant; mutation;
 KW D-N-carbamoyl-alpha-amino acid; penicillin; cephalosporin;
 KW production; racemisation.
 XX Bacillus thermoglucodasius.
 OS DE4328829-A.
 PN 02-MAR-1995.
 XX 27-AUG-1993; 93DE-4328829.
 XX 27-AUG-1993; 93DE-4328829.
 XX (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA Burtscher H, Lang G, Popp F;
 PI WPI; 1995-099331/14.
 XX N-PSDB: AAQ80565.
 XX New highly active and stable D-hydantoinase enzyme - isolated
 PT from Bacillus thermoglucodasius, useful in the prodn. of
 PT D-N-carbamoyl-alpha-amino acid for synthesis of, e.g. penicillin
 XX Claim 1; Page 5-6; 8pp; German.
 PS
 XX

CC The wild-type D-hydantoinase gene was isolated from *Bacillus*
 CC *thermoglucosaidus* using the amplification primers Hyd1 and Hyd2 (see
 CC AA080566 and AA080567). A single-base deletion in the HindIII site (from
 CC AAGCTT to AAGCT) was then introduced to give a coding region (AA080565)
 CC that encoded a D-hydantoinase (AA080565) that was shorter than the
 CC wild-type protein with a different C-terminal sequence (see Features
 CC Table). The mutant enzyme was found to have improved heat stability
 CC and higher activity compared to known hydantoinases.
 XX
 SQ Sequence 460 AA;

Query Match 1.7%; Score 10; DB 16; Length 460;
 Best Local Similarity 100.0%; Pred. No. 0.34;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 AVGSADLVI 411
 |||||
 Db 386 avgsadlvi 395

RESULT 11

AAW03025
 ID AAW03025 standard; Protein; 485 AA.

AC AAW03025;

DT 31-JAN-1997 (first entry)

DE Pseudomonas hydantoinase.

KW Hydantoinase; hydantoin; d-N-carbamoyl-alpha-amino acid;
 KW intermediate; drug synthesis; penicillin; cephalosporin;
 KW antibiotic.

OS Pseudomonas sp.

XX WO9620275-A1.

XX 04-JUL-1996.

PF 26-DEC-1995; 95WO-JP02688.

XX 28-DEC-1994; 94JP-0326865.

XX (KANF) KANEGAFUCHI KAGAKU KOGYO KK.

PI Ikenaka Y, Nanba H, Takahashi S, Takano M, Yajima K;
 PI Yamada Y;

DR WPI; 1996-321848/32.

DR N-PSDB; AAT31258.

PT Prodn. of D-N-carbamoyl amino acid from 5-substd. hydantoin - using
 PT a recombinant hydantoinase derived from a strain of *Pseudomonas*,
 PT *Agrobacterium* or *Bacillus*

PS Claim 19; Page 36-40; 54pp; English.

CC D-N-carbamoyl-alpha-amino acid is produced from a 5-substituted
 CC hydantoin by treatment with a hydantoinase expressed by a
 CC transformant microorganism carrying a vector containing DNA coding
 CC for the hydantoinase and derived from *Bacillus* sp. KKK245,
 CC *Agrobacterium* sp. KKK712 or *Pseudomonas* sp. KKK003A. The D-N-
 CC carbamoyl-alpha-amino acid can be used for the production of
 CC optically active alpha amino acids (especially D-phenylglycine and D-
 CC p-hydroxyphenylglycine) as intermediates for drug synthesis,
 CC especially for the production of semi-synthetic penicillin and
 CC cephalosporin antibiotics.

XX Sequence 485 AA;

Query Match 1.7%; Score 10; DB 17; Length 485;
 Best Local Similarity 100.0%; Pred. No. 0.35;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 NEFVAVTSTN 387
 |||||
 Db 358 nefvavtstn 367

RESULT 12

AAG46570

ID AAG46570 standard; Protein; 576 AA.

AC AAG46570;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 58602.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 576;

Best Loca. Similarity 100.0%; Pred. No. 0.41;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 214 ELGITGPEGH 223

Db 135 e.gitgpegh 144

RESULT 13
AAG46569
ID AAG46569 standard; Protein; 588 AA.
XX AAG46569;
AC AAG46569;
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58601.
DE Arabidopsis thaliana.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX
XX PD. 06-SEP-2000.
XX
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

```
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 588;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ELGIGTPEGH 223
Db 147 elgigtpegh 156

RESULT 14
AAG46568
ID AAG46568 standard; Protein; 683 AA.
XX
AC AAG46568;
```

```
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58600.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EPI033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
```

```
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145080.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148555.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.7%; Score 10; DB 21; Length 683;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 214 ELGITGPEGH 223
Db 242 elgitgpegh 251

RESULT 15
AAG23636
ID AAG23636 standard; Protein; 283 AA.
XX
AC AAG23636;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27018.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
```

KW termination sequence.

XX OS Arabidopsis thaliana.	PR 28-JUN-1999;	99US-0140823.
XX PN EP1033405-A2.	PR 29-JUN-1999;	99US-0140991.
XX PD 06-SEP-2000.	PR 30-JUN-1999;	99US-0141287.
XX PF 25-FEB-2000; 2000EP-0301439.	PR 01-JUL-1999;	99US-0141842.
XX PR 25-FEB-1999;	PR 02-JUL-1999;	99US-0142154.
PR 05-MAR-1999;	PR 06-JUL-1999;	99US-0142055.
PR 09-MAR-1999;	PR 08-JUL-1999;	99US-0142390.
PR 23-MAR-1999;	PR 09-JUL-1999;	99US-0142803.
PR 25-MAR-1999;	PR 12-JUL-1999;	99US-0142920.
PR 29-MAR-1999;	PR 13-JUL-1999;	99US-0142977.
PR 01-APR-1999;	PR 14-JUL-1999;	99US-0143542.
PR 06-APR-1999;	PR 15-JUL-1999;	99US-0143624.
PR 08-APR-1999;	PR 16-JUL-1999;	99US-0144005.
PR 16-APR-1999;	PR 19-JUL-1999;	99US-0144085.
PR 19-APR-1999;	PR 19-JUL-1999;	99US-0144325.
PR 21-APR-1999;	PR 19-JUL-1999;	99US-0144331.
PR 23-APR-1999;	PR 19-JUL-1999;	99US-0144332.
PR 28-APR-1999;	PR 19-JUL-1999;	99US-0144333.
PR 30-APR-1999;	PR 20-JUL-1999;	99US-0144334.
PR 04-MAY-1999;	PR 20-JUL-1999;	99US-0144352.
PR 05-MAY-1999;	PR 21-JUL-1999;	99US-0144884.
PR 06-MAY-1999;	PR 21-JUL-1999;	99US-0144814.
PR 07-MAY-1999;	PR 21-JUL-1999;	99US-0145086.
PR 11-MAY-1999;	PR 22-JUL-1999;	99US-0145088.
PR 14-MAY-1999;	PR 22-JUL-1999;	99US-0145087.
PR 14-MAY-1999;	PR 22-JUL-1999;	99US-0145089.
PR 14-MAY-1999;	PR 23-JUL-1999;	99US-0145192.
PR 14-MAY-1999;	PR 23-JUL-1999;	99US-0145145.
PR 14-MAY-1999;	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	PR 26-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	PR 27-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	PR 27-JUL-1999;	99US-0145918.
PR 14-MAY-1999;	PR 28-JUL-1999;	99US-0145919.
PR 14-MAY-1999;	PR 28-JUL-1999;	99US-0145951.
PR 14-MAY-1999;	PR 02-AUG-1999;	99US-0146386.
PR 14-MAY-1999;	PR 02-AUG-1999;	99US-0146388.
PR 14-MAY-1999;	PR 02-AUG-1999;	99US-0146389.
PR 14-MAY-1999;	PR 03-AUG-1999;	99US-0147038.
PR 14-MAY-1999;	PR 04-AUG-1999;	99US-0147204.
PR 14-MAY-1999;	PR 04-AUG-1999;	99US-0147302.
PR 14-MAY-1999;	PR 05-AUG-1999;	99US-0147260.
PR 14-MAY-1999;	PR 06-AUG-1999;	99US-0147303.
PR 14-MAY-1999;	PR 06-AUG-1999;	99US-0147416.
PR 14-MAY-1999;	PR 09-AUG-1999;	99US-0147493.
PR 14-MAY-1999;	PR 09-AUG-1999;	99US-0147935.
PR 14-MAY-1999;	PR 10-AUG-1999;	99US-0148171.
PR 14-MAY-1999;	PR 11-AUG-1999;	99US-0148319.
PR 14-MAY-1999;	PR 12-AUG-1999;	99US-0148341.
PR 14-MAY-1999;	PR 13-AUG-1999;	99US-0148565.
PR 14-MAY-1999;	PR 13-AUG-1999;	99US-0148684.
PR 14-MAY-1999;	PR 16-AUG-1999;	99US-0149368.
PR 14-MAY-1999;	PR 17-AUG-1999;	99US-0149175.
PR 14-MAY-1999;	PR 18-AUG-1999;	99US-0149426.
PR 14-MAY-1999;	PR 20-AUG-1999;	99US-0149722.
PR 14-MAY-1999;	PR 20-AUG-1999;	99US-0149723.
PR 14-MAY-1999;	PR 20-AUG-1999;	99US-0149929.
PR 14-MAY-1999;	PR 23-AUG-1999;	99US-0149902.
PR 14-MAY-1999;	PR 23-AUG-1999;	99US-0149930.
PR 14-MAY-1999;	PR 25-AUG-1999;	99US-0150566.
PR 14-MAY-1999;	PR 26-AUG-1999;	99US-0150884.
PR 14-MAY-1999;	PR 27-AUG-1999;	99US-0151065.
PR 14-MAY-1999;	PR 27-AUG-1999;	99US-0151066.
PR 14-MAY-1999;	PR 27-AUG-1999;	99US-0151066.
PR 14-MAY-1999;	PR 27-AUG-1999;	99US-0151080.
PR 14-MAY-1999;	PR 30-AUG-1999;	99US-0151303.
PR 14-MAY-1999;	PR 31-AUG-1999;	99US-0151438.
PR 14-MAY-1999;	PR 01-SEP-1999;	99US-0151930.
PR 14-MAY-1999;	PR 07-SEP-1999;	99US-0152363.

XX WO200058473-A2.
 XX
 XX
 PD 05-OCT-2000.
 XX
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX
 PI Shimkets RA, Leach M;
 XX
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC77230.
 XX

Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 4756-4757; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 169 AA;

Query Match 1.4%; Score 8; DB 21; Length 169;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 VEYNIFEG 436
 |||||
 Db 21 veynifeg 28

RESULT 18
 AAR82837
 ID AAR82837 standard; Protein; 457 AA.

XX AAR82837;
 AC
 XX 13-MAR-1996 (first entry)
 DT

DE Agrobacterium radiobacter D-hydantoinase.

XX D-hydantoinase; D-N-carbamylase; enzyme; stereospecific reaction;
 KW D-amino acid.
 XX

OS Agrobacterium radiobacter.

XX EP677585-A1.
 XX
 XX
 PD 18-OCT-1995.
 XX
 XX
 PF 24-MAR-1995; 95EP-0104393.
 XX
 XX
 PR 15-APR-1994; 94IT-0MI0726.
 XX
 XX
 PA (ENIE) ENIRICERCH SPA.
 XX
 XX
 PI Frascotti G, Galli G, Grandi G, Grifantini R;
 XX
 XX
 DR WPI; 1995-352764/46.
 DR N-PSDB; AAT01498.
 XX

Prod'n. of D-alpha amino acids from racemic 5-substd. hydantoin cpds.
 PT - using microorganisms contg. hydantoinase and carbamylase genes.

XX Disclosure; Fig.5A-5C; 44pp; English.

XX A. radiobacter is the donor microorganism for genes encoding
 CC D-hydantoinase and D-N-carbamylase which are expressed in
 CC Escherichia coli or Bacillus subtilis using plasmid pSM651. The
 CC resulting recombinant E. coli may be used to catalyze the
 CC stereospecific preparation of D-amino acids from racemic
 CC 5-substituted hydantoin compounds.

XX Sequence 457 AA;

Query Match 1.4%; Score 8; DB 16; Length 457;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 VAVGSDAD 408
 |||||
 Db 382 vavgsdad 389

RESULT 19
 AAR60325
 ID AAR60325 standard; Protein; 526 AA.

XX AAR60325;
 AC
 XX 01-MAR-1995 (first entry)
 DT

XX Streptomyces cholesterol oxidase.

XX Cholesterol oxidase; Streptomyces; enzyme; thermostability;
 KW diagnosis; endocrine disease; metabolism.

XX Streptomyces sp. SA-COO.

XX JP06169765-A.

XX 21-JUN-1994.

XX 05-JUL-1993; 93JP-0165558.

XX 05-OCT-1992; 92JP-0266130.

XX (TOYM) TOYOBO KK.

XX WPI; 1994-237584/29.

XX N-PSDB; AAQ70418.

XX Cholesterol oxidase with high thermostability - useful for the
 PT diagnosis of endocrine diseases

XX Claim 4; Page 11-13; 15pp; Japanese.

XX

CC Cholesterol oxidase is used for measurement of cholesterol which is
 CC an index for the diagnosis of endocrine diseases and abnormal
 CC metabolism. The enzyme is obtained by adding amino acid residues to
 CC the N-terminus of cholesterol oxidase derived from Streptomyces sp.
 CC SA-COO, which shows the same activity as that of the naturally
 CC occurring enzyme, but has higher thermostability. Cholesterol
 CC oxidase DNA sequences derived from Streptomyces sp. are given in
 CC AAQ70417-18.
 XX
 SQ Sequence 526 AA;

Query Match 1.4%; Score 8; DB 15; Length 526;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 AALAGGTT 102
 |||||
 Db 1 aalaggtt 8

RESULT 20
 AAW08932
 ID AAW08932 standard; Protein; 546 AA.

XX AC AAW08932;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Val121Ala.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

FH Key Location/Qualifiers
 FT Misc-difference 121
 FT /label= Val121Ala

XX JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOB0 KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 PT stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 9; Page 11-12; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 CC oxidases. These modified cholesterol oxidases may be prepared by
 CC mutation of the wild type sequence (see also AAW08930). These mutant
 CC cholesterol oxidases pref. have amino acid substitutions at positions
 CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
 CC than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 AALAGGTT 102
 |||||

Db 21 aalaggtt 28

RESULT 21
 AAW08933

ID AAW08933 standard; Protein; 546 AA.

XX AC AAW08933;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Arg135His.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

FH Key Location/Qualifiers
 FT Misc-difference 135
 FT /label= Arg135His

XX JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOB0 KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 PT stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 11; Page 13-14; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 CC oxidases. These modified cholesterol oxidases may be prepared by
 CC mutation of the wild type sequence (see also AAW08930). These mutant
 CC cholesterol oxidases pref. have amino acid substitutions at positions
 CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
 CC than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 AALAGGTT 102
 |||||
 Db 21 aalaggtt 28

RESULT 22
 AAW08934

ID AAW08934 standard; Protein; 546 AA.

XX AC AAW08934;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Val145Glu.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX FH Key Location/Qualifiers

FT Misc-difference 145
XX /label= Vall45Glu
PN JP08242860-A.
XX
PD 24-SEP-1996.
XX
PF 07-MAR-1995; 95JP-0047339.
XX
PR 07-MAR-1995; 95JP-0047339.
XX
PA (TOYM) TOYOBOKK.
XX
DR WPI; 1996-479904/48.
XX
PT Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
PS Claim 13; Page 14-15; 23pp; Japanese.
XX
CC The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
DB 21 aalaggtt 28
|||||

RESULT 23
AAW08935
ID AAW08935 standard; Protein; 546 AA.
XX
AC AAW08935;
XX
DT 18-FEB-1997 (first entry)
XX
DE Cholesterol oxidase, Ser103Thr, Vall45Glu.
XX
KW Cholesterol oxidase; mutation; thermal stability; substitution.
XX
OS Streptomyces sp. SA-COO.
XX
PH Key Location/Qualifiers
FT Misc-difference 103
FT /label= Ser103Thr
FT Misc-difference 121
FT /label= Vall21Ala
FT Misc-difference 145
FT /label= Vall45Glu
XX
PN JP08242860-A.
XX
PD 24-SEP-1996.
XX
PF 07-MAR-1995; 95JP-0047339.
XX
PR 07-MAR-1995; 95JP-0047339.
XX
PA (TOYM) TOYOBOKK.
XX
DR WPI; 1996-479904/48.
XX
PT Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
PS Claim 17; Page 17-18; 23pp; Japanese.
XX
CC The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
DB 21 aalaggtt 28
|||||

RESULT 23
AAW08935
ID AAW08935 standard; Protein; 546 AA.
XX
AC AAW08935;
XX
DT 18-FEB-1997 (first entry)
XX
DE Cholesterol oxidase, Ser103Thr, Vall45Glu.
XX
KW Cholesterol oxidase; mutation; thermal stability; substitution.
XX
OS Streptomyces sp. SA-COO.
XX
PH Key Location/Qualifiers
FT Misc-difference 103
FT /label= Ser103Thr
FT Misc-difference 145
FT /label= Vall45Glu
XX
PN JP08242860-A.
XX
PD 24-SEP-1996.
XX
PF 07-MAR-1995; 95JP-0047339.
XX
PR 07-MAR-1995; 95JP-0047339.
XX
PA (TOYM) TOYOBOKK.
XX
DR WPI; 1996-479904/48.
XX
PT Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
PS Claim 17; Page 17-18; 23pp; Japanese.
XX
CC The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
SQ Sequence 546 AA;

PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
PS Claim 15; Page 16-17; 23pp; Japanese.
XX
CC The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
DB 21 aalaggtt 28
|||||

RESULT 24
AAW08936
ID AAW08936 standard; Protein; 546 AA.
XX
AC AAW08936;
XX
DT 18-FEB-1997 (first entry)
XX
DE Cholesterol oxidase, Ser103Thr, Vall21Ala, Vall45Glu.
XX
KW Cholesterol oxidase; mutation; thermal stability; substitution.
XX
OS Streptomyces sp. SA-COO.
XX
PH Key Location/Qualifiers
FT Misc-difference 103
FT /label= Ser103Thr
FT Misc-difference 121
FT /label= Vall21Ala
FT Misc-difference 145
FT /label= Vall45Glu
XX
PN JP08242860-A.
XX
PD 24-SEP-1996.
XX
PF 07-MAR-1995; 95JP-0047339.
XX
PR 07-MAR-1995; 95JP-0047339.
XX
PA (TOYM) TOYOBOKK.
XX
DR WPI; 1996-479904/48.
XX
PT Modified cholesterol oxidase enzyme - having higher thermal
PT stability than the parent enzyme, useful in cholesterol determin.
PT reagent
XX
PS Claim 17; Page 17-18; 23pp; Japanese.
XX
CC The sequences given in AAW08931-36 represent modified cholesterol
CC oxidases. These modified cholesterol oxidases may be prepared by
CC mutation of the wild type sequence (see also AAW08930). These mutant
CC cholesterol oxidases pref. have amino acid substitutions at positions
CC 103Ser, 121Val, 135Arg and/or 145Val They have higher thermal stability
CC than the parent enzyme.
XX
SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||||
 Db 21 aalaggtt 28

RESULT 25

AAW08930
 ID AAW08930 standard; Protein; 546 AA.

XX AC AAW08930;

XX DT 18-FEB-1997 (first entry)

XX DE Wild type cholesterol oxidase.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOBO KK.

XX DR WPI; 1996-479904/48.

XX DR N-PSDB; AAT49320.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 3; Page 8-9; 23pp; Japanese.

XX CC This sequence represents wild type cholesterol oxidase. Modified
 cholesterol oxidases may be prepared by mutation of this
 sequence. Mutant cholesterol oxidases pref. have amino acid
 substitutions at positions 103Ser, 121Val, 135Arg and/or 145Val (see
 also AAW08931-36). The modified enzymes have higher thermal stability
 than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||||
 Db 21 aalaggtt 28

RESULT 26

AAW08931
 ID AAW08931 standard; Protein; 546 AA.

XX AC AAW08931;

XX DT 18-FEB-1997 (first entry)

XX DE Cholesterol oxidase, Ser103Thr.

XX KW Cholesterol oxidase; mutation; thermal stability; substitution.

XX OS Streptomyces sp. SA-COO.
 XX FH Key Location/Qualifiers
 FT Misc-difference 103
 FT /label= Ser103Thr

XX PN JP08242860-A.

XX PD 24-SEP-1996.

XX PF 07-MAR-1995; 95JP-0047339.

XX PR 07-MAR-1995; 95JP-0047339.

XX PA (TOYM) TOYOBO KK.

XX DR WPI; 1996-479904/48.

XX PT Modified cholesterol oxidase enzyme - having higher thermal
 stability than the parent enzyme, useful in cholesterol determin.

XX PS Claim 7; Page 10-11; 23pp; Japanese.

XX CC The sequences given in AAW08931-36 represent modified cholesterol
 oxidases. These modified cholesterol oxidases may be prepared by
 mutation of the wild type sequence (see also AAW08930). These mutant
 cholesterol oxidases pref. have amino acid substitutions at positions
 103Ser, 121Val, 135Arg and/or 145Val. They have higher thermal stability
 than the parent enzyme.

XX SQ Sequence 546 AA;

Query Match 1.4%; Score 8; DB 17; Length 546;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||||
 Db 21 aalaggtt 28

RESULT 27

AAW75425

ID AAW75425 standard; Protein; 837 AA.

XX AC AAW75425;

XX DT 30-MAR-1999 (first entry)

XX DE Human aggrecan degrading metalloprotease 1.

XX KW Human; aggrecan degrading metalloprotease; cartilage; proteoglycan;
 interglobular domain; matrix metalloprotease; bovine; interleukin-1beta;
 primer; PCR; amplification; inhibitor; cleavage; inhibitor; ADMP;
 arthritis; joint injury; pseudogout.

XX OS Homo sapiens.

XX PN WO9905291-A2.

XX PD 04-FEB-1999.

XX PF 24-JUL-1998; 98WO-US15438.

XX PR 16-OCT-1997; 97US-0062169.

XX PR 25-JUL-1997; 97US-0053850.

XX PR 15-AUG-1997; 97US-0055836.

XX PA (DUPO) DU PONT PHARM CO.

XX XX

PI Arner EC, Burn TC, Copeland RA, Decicco CP, Liu R, Magolda R;
 PI Pratta M, Solomon KA, Tortorella MD, Trzaskos JM, Yang F;
 DR WPI; 1999-142943/12.
 DR N-PSDB; AAX00725.
 XX
 XX
 PT New isolated aggrecan degrading metallo proteases - used to develop
 PT products for treating, e.g. Osteoarthritis, joint injury, reactive
 PT arthritis, psoriatic arthritis or juvenile rheumatoid arthritis
 XX
 XX Claim 25; page 61-62; 73pp; English.
 XX
 CC This sequence represents the human aggrecan degrading metalloprotease 1
 CC (ADMP-1). ADMP-1 and ADMP-2 (AAW75426) are novel proteases that cleave
 CC the aggrecan (a major cartilage proteoglycan) between residues
 CC Glu373-Ala374 of the interglobular domain (compared with cleavage between
 CC Asn341-Phe342 by the matrix metalloproteases MMP-1, -2, -3, -7, -8, -9
 CC and -13). ADMP-1 and ADMP-2 were isolated and purified from the
 CC conditioned media of bovine nasal cartilage stimulated by
 CC interleukin-1beta. The purified proteins were partially sequenced and
 CC primers were synthesised based on the resultant amino acid sequences
 CC (AAX00727-X00732 for ADMP-1 and AAX00733-X00736 for ADMP-2). These were
 CC used to isolate the corresponding genes from cDNA. The ADMP polypeptides
 CC can be used for identifying inhibitors of ADMP activity which would
 CC prevent cleavage of the aggrecan core protein, thereby decreasing the
 CC loss of aggrecan from cartilage. Such inhibitors can be used for treating
 CC diseases such as osteoarthritis, joint injury, reactive arthritis, acute
 CC pyrophosphate arthritis (pseudogout), psoriatic arthritis and juvenile
 CC rheumatoid arthritis.
 XX
 SQ Sequence 837 AA;

Query Match 1.4%; Score 8; DB 20; Length 837;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 PGSGAPAR 519
 Db 73 pgsgapar 80
 |||||
 |||||

RESULT 28
 AAY99429
 ID AAY99429 standard; Protein; 837 AA.
 XX
 AC AAY99429;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1563 (UNQ769) amino acid sequence SEQ ID NO:317.
 XX
 DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 XX
 OS Homo sapiens.
 XX
 OS WO200012708-A2.
 PN
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20111.
 XX
 XX 01-SEP-1998; 98US-0098716.
 PR 01-SEP-1998; 98US-0098749.
 PR 01-SEP-1998; 98US-0098750.
 PR 02-SEP-1998; 98US-0098803.
 PR 02-SEP-1998; 98US-0098821.
 PR 02-SEP-1998; 98US-0098843.
 PR 09-SEP-1998; 98US-0099536.
 PR 09-SEP-1998; 98US-0099596.
 PR 09-SEP-1998; 98US-0099598.
 PR 09-SEP-1998; 98US-0099602.

PR 09-SEP-1998; 98US-0099642.
 PR 10-SEP-1998; 98US-0099741.
 PR 10-SEP-1998; 98US-0099754.
 PR 10-SEP-1998; 98US-0099763.
 PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.
 PR 10-SEP-1998; 98US-0099815.
 PR 10-SEP-1998; 98US-0099816.
 PR 15-SEP-1998; 98US-0100385.
 PR 15-SEP-1998; 98US-0100388.
 PR 15-SEP-1998; 98US-0100390.
 PR 16-SEP-1998; 98US-0100584.
 PR 16-SEP-1998; 98US-0100627.
 PR 16-SEP-1998; 98US-0100661.
 PR 16-SEP-1998; 98US-0100662.
 PR 16-SEP-1998; 98US-0100664.
 PR 17-SEP-1998; 98US-0100683.
 PR 17-SEP-1998; 98US-0100684.
 PR 17-SEP-1998; 98US-0100710.
 PR 17-SEP-1998; 98US-0100711.
 PR 17-SEP-1998; 98US-0100930.
 PR 18-SEP-1998; 98US-0100930.
 PR 18-SEP-1998; 98US-0100848.
 PR 18-SEP-1998; 98US-0100849.
 PR 18-SEP-1998; 98US-0101014.
 PR 18-SEP-1998; 98US-0101068.
 PR 18-SEP-1998; 98US-0101071.
 PR 22-SEP-1998; 98US-0101279.
 PR 23-SEP-1998; 98US-0101471.
 PR 23-SEP-1998; 98US-0101472.
 PR 23-SEP-1998; 98US-0101474.
 PR 23-SEP-1998; 98US-0101475.
 PR 23-SEP-1998; 98US-0101476.
 PR 23-SEP-1998; 98US-0101477.
 PR 23-SEP-1998; 98US-0101479.
 PR 24-SEP-1998; 98US-0101738.
 PR 24-SEP-1998; 98US-0101741.
 PR 24-SEP-1998; 98US-0101743.
 PR 24-SEP-1998; 98US-0101915.
 PR 24-SEP-1998; 98US-0101916.
 PR 29-SEP-1998; 98US-0102207.
 PR 29-SEP-1998; 98US-0102240.
 PR 29-SEP-1998; 98US-0102307.
 PR 29-SEP-1998; 98US-0102330.
 PR 29-SEP-1998; 98US-0102331.
 PR 30-SEP-1998; 98US-0102484.
 PR 30-SEP-1998; 98US-0102487.
 PR 30-SEP-1998; 98US-0102570.
 PR 30-SEP-1998; 98US-0102571.
 PR 01-OCT-1998; 98US-0102684.
 PR 01-OCT-1998; 98US-0102687.
 PR 02-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0103258.
 PR 06-OCT-1998; 98US-0103449.
 PR 07-OCT-1998; 98US-0103314.
 PR 07-OCT-1998; 98US-0103315.
 PR 07-OCT-1998; 98US-0103328.
 PR 07-OCT-1998; 98US-0103395.
 PR 07-OCT-1998; 98US-0103396.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103633.
 PR 08-OCT-1998; 98US-0103678.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.

PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PI WPI; 2000-237871/20.
XX N-PSDB; AAA37111.
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 180; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridization probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 837 AA;
Query Match 1.4%; Score 8; DB 21; Length 837;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 512 PGSGAPAR 519
Db 73 pgsgapar 80
RESULT 30
AAW10617
ID AAW10617 standard; Protein; 15 AA.
XX
AC AAW106.7;

RESULT 29
AAB66178
ID AAB66173 standard; protein; 837 AA.
XX
XX AAB66173;
AC
XX 02-APR-2001 (first entry)
DT
XX Protein of the invention #90.
DE
XX Secreted; transmembrane; gene therapy.
KW
XX Unidentified.
OS
XX WO200078961-Al.
PN
XX 28-DEC-2000.
PD
XX 18-FEB-2000; 2000WO-US04342.
PF
XX 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH) GENENTECH INC.
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX
XX WPI; 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX
XX Claim 1; Fig 180; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 837 AA;
Query Match 1.4%; Score 8; DB 22; Length 837;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 512 PGSGAPAR 519
Db 73 pgsgapar 80
RESULT 30
AAW10617
ID AAW10617 standard; Protein; 15 AA.
XX
AC AAW106.7;

XX DT 29-OCT-1997 (first entry)
 XX DE L-galactono-1,4-lactone:ferricytochrome-c-oxidoreductase fragment 5.
 XX KW L-galactono-gamma-lactone; ascorbic acid; cauliflower; mitochondrion;
 XX KW antioxidant; chiral; L-galactonate; D-galacturonic acid.
 XX OS Brassica oleracea botrytis.
 XX PN WO9704100-A2.
 XX PD 06-FEB-1997.
 XX PF 17-JUL-1996; 96WO-DE01339.
 XX PR 17-JUL-1995; 95DE-4025991.
 XX PA (FRAU) FRAUNHOFER GES FOERDERUNG ANGEWANDTEN.
 XX PI Wissler J;
 XX DR N-PSDB; AAT61001.
 XX PS Claim 5; Page 26; 34pp; German.
 XX CC A new L-galactono-1,4-lactone:ferricytochrome-c-oxidoreductase
 CC enzyme was isolated from cauliflower mitochondria. The enzyme
 CC oxidises L-galactono-gamma-lactone to ascorbic acid, has an
 CC apparent molecular weight of 56 kD and includes the amino acid
 CC sequence: Val-Gln-Gln-Leu-Val-Asp-Ala-Ile-Gln-Glu-Tyr-Gly-Leu.
 CC A further 15 partial fragments of the enzyme have been obtained
 CC and sequenced, including the present sequence. The enzyme is used
 CC for ascorbic acid synthesis, particularly from waste materials,
 CC and for the determination of chiral L-galactonate or D-galacturonic
 CC acid. Ascorbic acid is well known as a nutritional supplement, free
 CC radical scavenger and food antioxidant.
 XX SQ Sequence 15 AA;

Query Match 1.2%; Score 7; DB 18; Length 15;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 KEELEAL 155
 Db 7 keeleal 13
 |||||

RESULT 31
 AAW69340
 ID AAW69340 standard; Protein; 55 AA.
 XX AC AAW69340;
 XX DT 25-NOV-1998 (first entry)
 XX DE Secreted protein of clone CD311_2.
 XX KW Secreted protein; nutritional activity; immune stimulant;
 KW immune suppressor; haematopoiesis regulator; tissue growth activity;
 KW activin/inhibin activity; chemotactic/chemokinetic activity;
 KW haemostatic activity; thrombolytic activity; receptor/ligand activity;
 KW anti-inflammatory activity; cadherin suppressor; tumour inhibitor;
 KW tumour invasion suppressor; therapy.
 XX OS Homo sapiens.

XX PN WO9837094-A2.
 XX PD 27-AUG-1998.
 XX PF 24-FEB-1998; 98WO-US03595.
 XX PR 23-FEB-1998; 98US-0028168.
 XX PR 24-FEB-1997; 97US-0804561.
 XX PA (GEMY) GENETICS INST INC.
 XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX DR WPI; 1998-467491/40.
 XX DR N-PSDB; AAV58536.
 XX PT Secreted proteins with biological activity - and encoding
 PT polynucleotide(s), useful e.g. in therapy and diagnosis of medical
 PT conditions and to identify protein agonists or antagonists
 XX PS Claim 15; Page 69; 111pp; English.
 XX CC This sequence is a secreted protein of the invention. This sequence
 CC is encoded by the DNA of clone CD311_2, which was isolated from a human
 CC foetal brain cDNA library. The polypeptides are predicted to have
 CC useful biological activities which would make them suitable for treating,
 CC preventing or ameliorating medical conditions in humans and animals,
 CC although no supporting biological data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity or other activities. They may also be used for diagnostic
 CC purposes. The polynucleotides are useful to produce the proteins or
 CC generate probes or primers to identify and/or amplify similar genes
 CC e.g. species homologues. They are also useful for gene therapy and to
 CC produce transgenic animals with altered gene expression.
 XX SQ Sequence 55 AA;

Query Match 1.2%; Score 7; DB 19; Length 55;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 539 FSLSGSQ 545
 Db 43 fslsgsq 49
 |||||

RESULT 32
 AAG33074
 ID AAG33074 standard; Protein; 90 AA.
 XX AC AAG33074;
 XX DT 18-OCT-2000 (first entry)
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 40020.
 XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.
 XX PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0128845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.


```
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 VTGSABC 334
Db 30 vtgsabc 36

RESULT 33
AAG48928
ID AAG48928 standard; Protein; 90 AA.
XX AC AAG48928;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61841.
XX DX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134250.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
```

```
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match. 1.2% Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 328 VTGSAHC 334
Db 30 vtgsahc 36
|||||

RESULT 34
AAG33073
ID AAG33073 standard; Protein; 91 AA.
XX
AC AAG33073;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 40019.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
```

PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0133256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	28-JUL-1999;	99US-0146386.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	02-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	03-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	04-AUG-1999;	99US-0147302.
PR	04-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147192.
PR	07-JUN-1999;	99US-0137724.	PR	05-AUG-1999;	99US-0147260.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147303.
PR	10-JUN-1999;	99US-0138540.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147493.
PR	14-JUN-1999;	99US-0139119.	PR	09-AUG-1999;	99US-0147935.
PR	16-JUN-1999;	99US-0139452.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139453.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139492.	PR	12-AUG-1999;	99US-0148341.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139455.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139456.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139457.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139458.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139461.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139462.	PR	20-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139463.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139750.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139763.	PR	25-AUG-1999;	99US-0150566.
PR	21-JUN-1999;	99US-0139817.	PR	25-AUG-1999;	99US-0150884.
PR	22-JUN-1999;	99US-0139899.	PR	26-AUG-1999;	99US-0151065.
PR	23-JUN-1999;	99US-0140353.	PR	27-AUG-1999;	99US-0151066.
PR	23-JUN-1999;	99US-0140354.	PR	27-AUG-1999;	99US-0151066.
PR	24-JUN-1999;	99US-0140695.	PR	27-AUG-1999;	99US-0151080.
PR	28-JUN-1999;	99US-0140823.	PR	30-AUG-1999;	99US-0151303.
PR	29-JUN-1999;	99US-0140991.	PR	31-AUG-1999;	99US-0151438.
PR	30-JUN-1999;	99US-0141287.	PR	01-SEP-1999;	9

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2% Score 7; DB 21; Length 91;

Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 VTGSAHC 334
Db 31 vtgsahc 37

RESULT 35

AAG48927
ID AAG48927 standard; Protein; 91 AA.

XX AAG48927;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 61840.

KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151330.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159320.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161994.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2% Score 7; DB 21; Length 91;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 328 VTGSAHC 334
Db 31 vtgsahc 37
|||||

RESULT 36

AAB80092
ID AAB80092 standard; Protein; 93 AA.

AC AAB80092;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:918.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.

OS Corynebacterium glutamicum.

PN WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030476.

PR 02-JUL-1999; 99US-0142101.

PR 08-JUL-1999; 99DE-1031415.

PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031435.

PR 08-JUL-1999; 99DE-1031443.

PR 08-JUL-1999; 99DE-1031453.

PR 08-JUL-1999; 99DE-1031457.

PR 08-JUL-1999; 99DE-1031465.

PR 08-JUL-1999; 99DE-1031478.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031541.

PR 08-JUL-1999; 99DE-1031573.

PR 08-JUL-1999; 99DE-1031592.

PR 08-JUL-1999; 99DE-1031632.

PR 08-JUL-1999; 99DE-1031634.

PR 08-JUL-1999; 99DE-1031636.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 03-SEP-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX

(BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI: 2001-137957/14.

N-PSDB; AAF72211.

Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway proteins, useful for producing fine chemicals in microorganisms, including organic acids, nonproteinogenic amino acids, and purine and pyrimidine bases -

Claim 20; Page 1416; 1737pp; English.

AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum MP nucleic acids are useful for the production of fine chemicals in microorganisms, including organic acids, nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides and enzymes.

Sequence 93 AA;

Query Match 1.2%; Score 7; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 316 HLTCLLS 322

Db 10 hltclls 16

RESULT 37

AAW63722

ID AAW63722 standard; Protein; 110 AA.

XX

AC

AAW63722;

XX

DT

08-OCT-1998 (first entry)

XX

DE

C. histolyticum CHCI protein N-terminal.

XX

KW

Class I collagenase; CHCI; isolation; transplant; immunotherapy; ulcer;

XX

OS

identification; gene therapy; tumour.

XX

PN

Clostridium histolyticum.

XX

PN

WO9822574-A2.

XX

PD

28-MAY-1998.

XX

PF

30-OCT-1997; 97WO-EP06002.

XX

PR

24-JAN-1997; 97EP-0101085.

PR

19-NOV-1996; 96EP-0118490.

XX

PA

(BOEF) BOEHRINGER MANNHEIM GMBH.

XX

PI

Ambrosius D, Burtscher H, Hesse F;

XX

DR

WPI; 1998-312460/27.

XX

PT

New recombinant collagenase type I from Clostridium histolyticum -

PT

useful for isolating cells from animal tissue, e.g. for

PT

transplantation or immuno-therapy and in gene therapy

XX

PS

Claim 1; Page 27; 38pp; German.

XX

CC

This sequence represents the N-terminal from a Class I collagenase, CHCI from Clostridium histolyticum. This protein can be used to isolate cells from animal or human tissue, e.g. for use in transplantation or immunotherapy (particularly pancreatic islets). This protein may also be used for identification of rare cells carrying specific markers, for use in gene therapy of a wide variety of tissues, including solid tumours, and for cleaning ulcers. Recombinant CHCI has high enzymatic activity and is very stable. It can be produced very pure (free of other clostridial enzymes and toxins) and in high yield. Compositions containing the protein may be used where cells are to be recovered without alteration of their cell-surface antigens.

XX

SQ

Sequence 110 AA;

Query Match

Best Local Similarity 1.2%; Score 7; DB 19; Length 110;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 114 VSLLAAY 120

Db 28 vsllaay 34

RESULT 38

AAG24039

ID AAG24039 standard; Protein; 121 AA.

XX

AC

AAG24039;

XX

DT

17-OCT-2000 (first entry)

XX

DE

Arabidopsis thaliana protein fragment SEQ ID NO: 27563.

XX

KW

Protein identification; signal transduction pathway; metabolic pathway;

KW

hybridisation assay; genetic mapping; gene expression control; promoter;

XX

OS

Arabidopsis thaliana.

XX

PN

EP1033405-A2.

XX PD 06-SEP-2000. 99US-0142055.
XX PF 25-FEB-2000; 2000EP-0301439. PR 02-JUL-1999;
XX XX 99US-0142390. PR 06-JUL-1999;
XX 99US-0142803. PR 08-JUL-1999;
XX 99US-0142920. PR 09-JUL-1999;
XX 99US-0142977. PR 12-JUL-1999;
XX 99US-0143542. PR 13-JUL-1999;
XX 99US-0143624. PR 14-JUL-1999;
XX 99US-0144005. PR 15-JUL-1999;
XX 99US-0144085. PR 16-JUL-1999;
XX 99US-0144086. PR 16-JUL-1999;
XX 99US-0144325. PR 19-JUL-1999;
XX 99US-0144331. PR 19-JUL-1999;
XX 99US-0144332. PR 19-JUL-1999;
XX 99US-0144333. PR 19-JUL-1999;
XX 99US-0144334. PR 19-JUL-1999;
XX 99US-0144335. PR 19-JUL-1999;
XX 99US-0144352. PR 20-JUL-1999;
XX 99US-0144632. PR 20-JUL-1999;
XX 99US-0144884. PR 20-JUL-1999;
XX 99US-0144814. PR 21-JUL-1999;
XX 99US-0145086. PR 21-JUL-1999;
XX 99US-0145088. PR 21-JUL-1999;
XX 99US-0145085. PR 22-JUL-1999;
XX 99US-0145087. PR 22-JUL-1999;
XX 99US-0145089. PR 22-JUL-1999;
XX 99US-0145192. PR 22-JUL-1999;
XX 99US-0145145. PR 23-JUL-1999;
XX 99US-0145210. PR 23-JUL-1999;
XX 99US-0145224. PR 23-JUL-1999;
XX 99US-0145276. PR 26-JUL-1999;
XX 99US-0145913. PR 27-JUL-1999;
XX 99US-0145918. PR 27-JUL-1999;
XX 99US-0145919. PR 27-JUL-1999;
XX 99US-0145951. PR 28-JUL-1999;
XX 99US-0146386. PR 02-AUG-1999;
XX 99US-0146388. PR 02-AUG-1999;
XX 99US-0146389. PR 02-AUG-1999;
XX 99US-0147038. PR 03-AUG-1999;
XX 99US-0147204. PR 04-AUG-1999;
XX 99US-0147302. PR 04-AUG-1999;
XX 99US-0147192. PR 05-AUG-1999;
XX 99US-0147260. PR 05-AUG-1999;
XX 99US-0147303. PR 06-AUG-1999;
XX 99US-0147416. PR 06-AUG-1999;
XX 99US-0147493. PR 09-AUG-1999;
XX 99US-0147935. PR 09-AUG-1999;
XX 99US-0148171. PR 10-AUG-1999;
XX 99US-0148319. PR 11-AUG-1999;
XX 99US-0148341. PR 12-AUG-1999;
XX 99US-0148565. PR 13-AUG-1999;
XX 99US-0148684. PR 16-AUG-1999;
XX 99US-0149368. PR 16-AUG-1999;
XX 99US-0149175. PR 17-AUG-1999;
XX 99US-0149426. PR 18-AUG-1999;
XX 99US-0149722. PR 20-AUG-1999;
XX 99US-0149723. PR 20-AUG-1999;
XX 99US-0149929. PR 20-AUG-1999;
XX 99US-0149902. PR 23-AUG-1999;
XX 99US-0149930. PR 23-AUG-1999;
XX 99US-0150560. PR 25-AUG-1999;
XX 99US-0150884. PR 26-AUG-1999;
XX 99US-0151065. PR 27-AUG-1999;
XX 99US-0151066. PR 27-AUG-1999;
XX 99US-0151080. PR 27-AUG-1999;
XX 99US-0151303. PR 30-AUG-1999;
XX 99US-0151438. PR 31-AUG-1999;
XX 99US-0151930. PR 01-SEP-1999;
XX 99US-0152363. PR 07-SEP-1999;
XX 99US-0153070. PR 10-SEP-1999;
XX 99US-0153758. PR 13-SEP-1999;
XX 99US-0154018. PR 15-SEP-1999;
XX 99US-0154039. PR 16-SEP-1999;
XX 99US-0154779. PR 20-SEP-1999;

PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157665.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162143.

Query Match 1.2%; Score 7; DB 21; Length 121;

Best Local Similarity 100.0%; Pred. No. 91;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LEALYKE 158

Db 51 lealvke 57

RESULT 39

AAW20410
ID AAW20410 standard; protein; 124 AA.

AC AAW20410;

DT 29-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 3261306.aa.

KW Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope;
 KW cofactor metabolism.

OS Helicobacter pylori.

XX Key Location/Qualifiers

FT Misc-difference 72

FT /label= unknown

FT /note= "encoded by RGC codon"

FT Misc-difference 73

FT /note= "encoded by GTW codon"

XX

PN WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI; 1997-052306/05.

XX N-PSDB; AAT67792.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

XX Claim 61; Page 591; 1481pp; English.

XX The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in cofactor metabolism.

CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX Sequence 124 AA;

Query Match 1.2%; Score 7; DB 18; Length 124;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 VFGEPT 280

Db 39 vfgept 45

RESULT 40

AAW24654

ID AAW24654 standard; protein; 124 AA.

XX AAW24654;

XX 11-AUG-1997 (first entry)

DE H. pylori cytoplasmic protein 3261306.aa.

KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
 KW detection; antisense; inhibition.

OS Helicobacter pylori.

XX Key Location/Qualifiers

FT Misc-difference 72

FT /note= "encoded by RGC"

FT Misc-difference 73

FT /note= "encoded by GTW"

XX

PN WO9719098-A1.
 PD 29-MAY-1997.
 XX 15-NOV-1996; 96WO-US18542.
 XX 17-NOV-1995; 95US-0561469.
 PR (ASTR) ASTRA AB.
 XX Smith DH;
 PI WPI; 1997-298052/27.
 DR N-PSDB; AAT77472.
 XX Helicobacter pylori nucleic acid sequences and related proteins -
 used for diagnostics and therapeutics
 XX Claim 18; Page 173; 235pp; English.
 XX This sequence represents an H. pylori cytoplasmic protein involved in
 cofactor metabolism.
 CC Helicobacter pylori has been strongly linked to chronic gastritis and
 duodenal ulcer disease. The nucleic acid sequences of the invention
 are used to evaluate compounds, especially activators or inhibitors of
 bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 sequence. The nucleic acid sequences, and corresponding proteins, are
 also useful for generating vaccines for immunising subjects against H.
 pylori or for use in detecting the presence of Helicobacter species in
 a sample. Antisense nucleic acid sequences of these sequences are
 used to inhibit expression of a gene from Helicobacter species. H.
 pylori whole genomic DNA was isolated and nebulised to a median size of
 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 complementary to the BstXI-cut pMPX vectors, while the overhang is not
 self-complementary. Therefore the linkers will not concatamerise nor
 will the cut vector re-ligate itself easily. The linker-adaptor inserts
 were ligated to each of the 20 pMPX vectors to construct a series of
 shotgun subclone libraries. The purified DNA samples were then
 sequenced.
 CC Note: The ORF/protein reference number for this sequence was obtained
 from the related specification, WO9640893.
 XX Sequence 124 AA;
 SQ
 Query Match 1.2%; Score 7; DB 18; Length 124;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 274 VFGEPIIT 280
 Db 39 vfgepit 45
 |||||
 RESULT 41
 AAR62918
 ID AAR62918 standard; protein; 125 AA.
 XX AC AAR62918;
 XX 18-JUL-1995 (first entry)
 XX Human cytomegalovirus antibody heavy chain variable region.
 XX Human cytomegalovirus; antibody heavy chain variable region;
 KW HMCV; immunoassay; immunotherapy.
 XX Homo sapiens.
 XX WO9425490-A.
 XX 10-NOV-1994.
 XX 29-APR-1994; 94WO-US04705.
 XX 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 XX Claim 1; Page 149; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.
 CC A monoclonal antibody (Mab) containing a pair of these regions can be
 used to detect HCMV, and anti-human HCMV Abs in human patients via
 a competitive immunoassay. The Mabs may also be useful in
 immunotherapy.
 CC Sequence 125 AA;
 SQ
 Query Match 1.2%; Score 7; DB 15; Length 125;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 537 SGFSLSG 543
 Db 21 sgfsisg 27
 |||||
 RESULT 42
 AAR62920
 ID AAR62920 standard; protein; 125 AA.
 XX AC AAR62920;
 XX 18-JUL-1995 (first entry)
 XX Human cytomegalovirus antibody heavy chain variable region.
 XX Human cytomegalovirus; antibody heavy chain variable region;
 KW HMCV; immunoassay; immunotherapy.
 XX Homo sapiens.
 XX WO9425490-A.
 XX 10-NOV-1994.
 XX 29-APR-1994; 94WO-US04705.
 XX 30-APR-1993; 93US-0055985.
 XX (SCRI) SCRIPPS RES INST.
 XX Barbas C, Burion R, Burton DR, Williamson A, Burioni R;
 WPI; 1994-358194/44.
 XX Human monoclonal antibodies (MABs) against human cytomegalovirus
 - also nucleic acids and cell lines producing the MABs, useful in
 diagnosis and immunotherapy
 XX Claim 1; Page 149; 171pp; English.
 XX AAR62618-R62920 and AAR62921-R62923 are human cytomegalovirus (HMCV)
 antibody heavy chain and light chain variable regions respectively.

CC A monoclonal antibody (Mab) containing a pair of these regions can be
 CC used to detect HCMV, and anti-human HCMV Abs in human patients via
 CC a competitive immunoassay. The Mabs may also be useful in
 CC immunotherapy.

XX SQ Sequence 125 AA;

Query Match 1.2%; Score 7; DB 15; Length 125;
 Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543

|||||

Db 21 sgfslsq 27

RESULT 43

AAR36700

ID AAR36700 standard; Protein: 126 AA.

XX AC AAR36700;

XX DT 20-AUG-1993 (first entry)

XX DE Flagellin (amino acids 137-262).

XX KW CMP-KDO synthetase; pB776; pB410; capture; immunodot;

XX KW assay; Treponema pallidum; syphilis;

XX KW CTP: CMP-3-deoxy-manno-octulosonate cytidyl transferase.

XX OS Borrelia burgdorferi.

XX PN WO9308208-A.

XX PD 29-APR-1993.

XX PF 21-OCT-1992; 92WO-US09199.

XX PR 21-OCT-1991; 91US-0779704.

XX PA (ABBO) ABBOTT LAB.

XX PI Hunt JC, Pilot-Matias TJ, Robinson JW;

XX DR WPI; 1993-152418/18.

XX PT N-PSDB; AAQ41495.

XX PS Synthetic polypeptide immuno-reactive with Borrelia burgdorferi

XX PT antibodies - for diagnosis of Lyme disease

XX PS Claim 1; Page 48-49; 76pp; English.

XX CC Expression vector pB410 encodes fusion protein CKS-410, consisting

CC of 239 amino acids of CKS, 11 amino acids contributed by linker DNA

CC sequences, and amino acids 137 to 262 of B. burgdorferi flagellin.

CC The recombinant fusion proteins p776 and p410 are differentiating

CC polypeptides which can be used in capture assays and immunodot

CC assays for the detection of B. burgdorferi infection such as in

CC Lyme disease, without false positives due to Treponema pallidum

CC infection such as in patients with syphilis.

XX SQ Sequence 126 AA;

Query Match 1.2%; Score 7; DB 14; Length 126;

Best Local Similarity 100.0%; Pred. No. 94;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546

|||||

Db 27 slsgsga 33

RESULT 44

AAG24038

ID AAG24038 standard; Protein: 134 AA.

XX AC AAG24038;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 27562.

XX KW protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 16-APR-1999; 99US-0128714.

XX PR 19-APR-1999; 99US-0129845.

XX PR 21-APR-1999; 99US-0130077.

XX PR 23-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 05-MAY-1999; 99US-0132484.

XX PR 06-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 14-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 10-JUN-1999; 99US-0138847.

XX PR 14-JUN-1999; 99US-0139119.

XX PR 16-JUN-1999; 99US-0139452.

XX PR 16-JUN-1999; 99US-0139453.

XX PR 17-JUN-1999; 99US-0139492.

XX PR 18-JUN-1999; 99US-0139454.

XX PR 18-JUN-1999; 99US-0139455.

XX PR 18-JUN-1999; 99US-0139456.

XX PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0144086.
PR 12-JUL-1999; 99US-0144325.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 7; DB 21; Length 134;
Best Local Similarity 100.0%; Pred. No. 99;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LEALVKE 158
|||||
Db 64 lealvke 70

RESULT 45

AA43865

ID AA43865 standard; Peptide; 139 AA.

XX

AC AA43865;

XX

```

DT 11-FEB-2000 (first entry)
XX Heavy chain (VH) gene sequence of Bacillus anthracis antibody 9-1.
DE
XX Heavy chain; VH; IgG; monoclonal antibody; spore; Bacillus;
KW VH gene usage; anthrax.
XX
OS Mus sp.
XX
XX WO9955842-A1.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-US09122.
XX
XX 29-APR-1998; 98US-0069628.
XX
XX (UABR-) UAB RES FOUND.
XX
XX Kearney JF;
XX
XX WPI; 2000-013435/01.
XX
XX Monoclonal antibody specific for Bacillus spores, used to detect
PT anthrax -
XX
XX Example 13; Page 52-53; 64pp; English.
XX
XX AAY43862-75 represent the amino acid sequences of the heavy chain
CC (VH) gene sequences from antibodies against Bacillus anthracis
CC spores. The antibodies are produced by different hybridomas. The
CC specification describes monoclonal antibodies (especially IgG
CC antibodies) which are highly specific and can discriminate between the
CC spores of the Bacillus family. The antibodies are produced by exposing
CC mice to Bacillus spores. The humoral immune response to Bacillus spores
CC shows a conservation of VH gene usage which is distinct for each spore.
CC Peptide fragments derived from the antibodies are also capable of
CC binding spores. The monoclonal antibody, and peptide fragments of it, can
CC be used to detect Bacillus spores in a field sample. It is particularly
XX uses for detecting anthrax in a field sample.
XX
SQ Sequence 139 AA;

Query Match 1.28; Score 7; DB 21; Length 139;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
DB 12 sgfsisg 18
|||||

Search completed: July 30, 2001, 11:42:55
Job time: 139 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:41:01 ; Search time 12.41 Seconds
(without alignments)
949.047 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFGKKSIPRITSRLIR.....RTAQKIMAPGGRSNTLSLS 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 197339 seqs, 20590346 residues

Word size : 0

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	10	1.7	460	1	US-08-289-709-1
2	10	1.7	460	1	US-08-602-656-1
3	7	1.2	126	1	US-08-500-222-8
4	7	1.2	126	1	US-08-500-125-8
5	7	1.2	126	2	US-07-779-704B-8
6	7	1.2	248	1	US-08-500-222-4
7	7	1.2	248	1	US-08-500-125-4
8	7	1.2	248	2	US-07-779-704B-4
9	7	1.2	336	1	US-07-641-143B-2
10	7	1.2	336	1	US-08-124-290-2
11	7	1.2	336	3	US-08-696-372A-2
12	7	1.2	336	4	US-08-235-836C-52
13	7	1.2	336	4	US-08-235-836C-54
14	7	1.2	341	4	US-09-196-293-14
15	7	1.2	341	4	US-08-209-603E-14
16	7	1.2	376	1	US-08-500-222-6
17	7	1.2	376	1	US-08-500-125-6
18	7	1.2	376	2	US-07-779-704B-6
19	7	1.2	464	1	US-07-688-352C-16
20	7	1.2	464	2	US-08-474-379C-16
21	7	1.2	464	3	US-09-146-249A-16
22	7	1.2	464	3	US-08-206-188B-16
23	7	1.2	464	5	PCT-US91-02714-16
24	7	1.2	498	1	US-08-500-222-2
25	7	1.2	498	1	US-08-500-125-2
26	7	1.2	498	2	US-07-779-704B-2
27	7	1.2	514	4	US-08-796-899-25

28 6 1.0 8 3 US-09-188-579-62 Sequence 62, Appl
29 6 1.0 8 4 US-08-960-780-69 Sequence 69, Appl
30 6 1.0 8 4 US-09-315-444-62 Sequence 62, Appl
31 6 1.0 8 4 US-09-073-898-69 Sequence 69, Appl
32 6 1.0 10 3 US-08-159-339A-1037 Sequence 1037, Ap
33 6 1.0 15 2 US-08-480-190-162 Sequence 162, App
34 6 1.0 15 2 US-08-488-379-162 Sequence 162, App
35 6 1.0 15 5 PCT-US93-07545-162 Sequence 162, App
36 6 1.0 17 2 US-08-480-190-161 Sequence 161, App
37 6 1.0 17 2 US-08-488-379-161 Sequence 161, App
38 6 1.0 17 5 PCT-US93-07545-161 Sequence 161, App
39 6 1.0 17 5 PCT-US95-08596-21 Sequence 21, Appl
40 6 1.0 17 5 PCT-US95-08621-20 Sequence 20, Appl
41 6 1.0 18 2 US-08-480-190-160 Sequence 160, App
42 6 1.0 18 2 US-08-488-379-160 Sequence 160, App
43 6 1.0 18 5 PCT-US93-07545-160 Sequence 160, App
44 6 1.0 20 1 US-08-468-543-10 Sequence 10, Appl
45 6 1.0 20 2 US-08-469-692-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-08-289-709-1

; Sequence 1, Application US/08289709

; Patent No 5523224

; GENERAL INFORMATION:

; APPLICANT: Burtischer, Helmut; Lang, Gunter; Popp, Friedrich

; TITLE OF INVENTION: Recombinant D-Hydantoinase, A Process for the

; TITLE OF INVENTION: Production and Use

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/289,709

; FILING DATE: 12-AUGUST-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: P 43 28 829.4

; FILING DATE: 27-AUGUST-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5523224man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: BOER 1041

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200

; TELEFAX: (212) 838-3884

; INFORMATION FOR SEQ ID NO: 1:

; LENGTH: 460 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-289-709-1

Query Match 1.7%; Score 10; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 AV3SDADIVI 411

|||||||

```
Db 386 AVGSDADLVI 395

RESULT 2
US-08-602-656-1
; Sequence 1, Application US/08602656
; Patent No. 5679571
; GENERAL INFORMATION:
; APPLICANT: Bartscher, Helmut; Lang, Gunter; Popp, Friedrich
; TITLE OF INVENTION: Recombinant D-Hydantoinease, A Process For the
; TITLE OF INVENTION: Production and Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602.656
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,709
; FILING DATE: 12-AUGUST-1994
; APPLICATION NUMBER: P 43 28 829.4
; FILING DATE: 27-AUGUST-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5679571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: BOER 1041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-602-656-1

Query Match 1.7%; Score 10; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 402 AVGSDADLVI 411
Db 386 AVGSDADLVI 395
|||||
|||||

RESULT 3
US-08-500-222-8
; Sequence 8, Application US/08500222
; Patent No. 5643733
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park

Qy 402 AVGSDADLVI 411
Db 386 AVGSDADLVI 395
|||||
|||||

Query Match 1.7%; Score 10; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 402 AVGSDADLVI 411
Db 386 AVGSDADLVI 395
|||||
|||||

RESULT 4
US-08-500-125-8
; Sequence 8, Application US/08500125
; Patent No. 5643751
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,125
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,704
; FILING DATE: 21-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; HYPOTHETICAL:
; ANTI-SENSE: NO
; FRAGMENT TYPE: Interval
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: B31
US-08-500-222-8

Query Match 1.2%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLGSGQA 546
Db 27 SLGSGQA 33
|||||
|||||

RESULT 4
US-08-500-125-8
; Sequence 8, Application US/08500125
; Patent No. 5643751
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,125
; FILING DATE:
; CLASSIFICATION: 435
```


;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/779,704
;; FILING DATE: 21-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wong, Wean Khing
;; REGISTRATION NUMBER: 33561
;; REFERENCE/DOCKET NUMBER: 5051.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 126 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: Protein
;; HYPOTHETICAL:
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: interval
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-08-500-125-8

Query Match 1.2%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
Db 27 SLGSGQA 33

RESULT 5
US-07-779-704B-8
;; Sequence 8, Application US/07779704B
;; Patent No. 5965702
;; GENERAL INFORMATION:
;; APPLICANT: Robinson, John M
;; APPLICANT: Pilot-Matias, Tami J
;; APPLICANT: Hunt, Jeffrey C
;; TITLE OF INVENTION: Borrelia burgdorferi antigens
;; TITLE OF INVENTION: and uses thereof
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Abbott Laboratories
;; STREET: One Abbott Park Road
;; CITY: Abbott Park
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60064
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/779,704B
;; FILING DATE: 19911021
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wong, Wean Khing
;; REGISTRATION NUMBER: 33561
;; REFERENCE/DOCKET NUMBER: 5051.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 126 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear

;; MOLECULE TYPE: Protein
;; HYPOTHETICAL:
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: interval
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-07-779-704B-8

Query Match 1.2%; Score 7; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
Db 27 SLGSGQA 33

RESULT 6
US-08-500-222-4
;; Sequence 4, Application US/08500222
;; Patent No. 5643733
;; GENERAL INFORMATION:
;; APPLICANT: Robinson, John M
;; APPLICANT: Pilot-Matias, Tami J
;; APPLICANT: Hunt, Jeffrey C
;; TITLE OF INVENTION: Borrelia burgdorferi antigens
;; TITLE OF INVENTION: and uses thereof
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Abbott Laboratories
;; STREET: One Abbott Park Road
;; CITY: Abbott Park
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60064
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/500,222
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/779,704
;; FILING DATE: 21-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wong, Wean Khing
;; REGISTRATION NUMBER: 33561
;; REFERENCE/DOCKET NUMBER: 5051.US.01
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 708-937-9396
;; TELEFAX: 708-937-9556
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 248 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: internal
;; ORIGINAL SOURCE:
;; ORGANISM: Borrelia burgdorferi
;; STRAIN: B31
;; US-08-500-222-4

Query Match 1.2%; Score 7; DB 1; Length 248;
Best Local Similarity 100.0%; Pred. No. 44;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 7

US-08-500-125-4
 ; Sequence 4, Application US/08500125
 ; Patent No. 5643751

; GENERAL INFORMATION:
 ; APPLICANT: Robinson, John M
 ; APPLICANT: Pilot-Matias, Tami J
 ; APPLICANT: Hunt, Jeffrey C
 ; TITLE OF INVENTION: Borrelia burgdorferi antigens
 ; TITLE OF INVENTION: and uses thereof
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: One Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/500.125
 ; FILING DATE: 21-OCT-1991

; CLASSIFICATION: 435
 ; PRIOR APPLICATION NUMBER: US 07/779,704
 ; FILING DATE: 21-OCT-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wong, Wean Khing
 ; REGISTRATION NUMBER: 33561
 ; REFERENCE/DOCKET NUMBER: 5051.US.01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 708-937-9396
 ; TELEFAX: 708-937-9536

; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 248 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Borrelia burgdorferi
 ; STRAIN: B31
 ; US-08-500-125-4

Query Match 1.2%; Score 7; DB 1; Length 248;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 8

US-07-779-704B-4
 ; Sequence 4, Application US/07779704B
 ; Patent No. 5965702
 ; GENERAL INFORMATION:

; APPLICANT: Robinson, John M
 ; APPLICANT: Pilot-Matias, Tami J
 ; APPLICANT: Hunt, Jeffrey C
 ; TITLE OF INVENTION: Borrelia burgdorferi antigens
 ; TITLE OF INVENTION: and uses thereof
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Abbott Laboratories
 ; STREET: One Abbott Park Road
 ; CITY: Abbott Park
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60064

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/779.704B
 ; FILING DATE: 19911021
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wong, Wean Khing
 ; REGISTRATION NUMBER: 33561
 ; REFERENCE/DOCKET NUMBER: 5051.US.01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 708-937-9396
 ; TELEFAX: 708-937-9536

; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 248 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Borrelia burgdorferi
 ; STRAIN: B31
 ; US-07-779-704B-4

Query Match 1.2%; Score 7; DB 2; Length 248;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLSGSQA 546
 Db 100 SLSGSQA 106

RESULT 9

US-07-641-143B-2
 ; Sequence 2, Application US/07641143B
 ; Patent No. 5436000
 ; GENERAL INFORMATION:

; APPLICANT: Barbour, Alan G.
 ; APPLICANT: Bundoc, Virgilio
 ; TITLE OF INVENTION: Flagella-less Borrelia
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/641,143B
;; FILING DATE: 11-JAN-1991
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Patterson, Melinda
;; REGISTRATION NUMBER: 33,062
;; REFERENCE/DOCKET NUMBER: UTSK092
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 713-787-1592
;; TELEFAX: 713-789-2679
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 336 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-07-641-143B-2

Query Match 1.2%; Score 7; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQ 546
Db 163 SLSGSQ 169

RESULT 10
US-08-124-290-2
; Sequence 2, Application US/08124290
; Patent No. 5585102
; GENERAL INFORMATION:
; APPLICANT: Barbour, Alan G.
; APPLICANT: Bundoc, Virgilio
; TITLE OF INVENTION: Flagella-less Borrelia
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/124,290
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/641,143
FILING DATE: 11-JAN-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Patterson, Melinda
REGISTRATION NUMBER: 33,062
REFERENCE/DOCKET NUMBER: UTSK092
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1592
TELEFAX: 713-789-2679
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-124-290-2

Query Match 1.2%; Score 7; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQ 546
Db 163 SLSGSQ 169

RESULT 11
US-08-696-372A-2
; Sequence 2, Application US/08696372A
; Patent No. 6077515
; GENERAL INFORMATION:
; APPLICANT: Barbour, Alan G.
; APPLICANT: Bundoc, Virgilio
; TITLE OF INVENTION: Flagella-less Borrelia
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,372A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/641,143
FILING DATE: 11-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Patterson, Melinda
REGISTRATION NUMBER: 33,062
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1592
TELEFAX: 713-789-2679
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-696-372A-2

Query Match 1.2%; Score 7; DB 3; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSQ 546
Db 163 SLSGSQ 169

RESULT 12
US-08-235-836C-52
; Sequence 52, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: NO. 6248562e1 Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Brookhaven National Laboratory
;; STREET:
;; CITY: Upton
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 11973
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/235.836C
;; FILING DATE: 29-APR-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/148,191
;; FILING DATE: 01-11-93
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bogosian, Margaret C.
;; REGISTRATION NUMBER: 25,324
;; REFERENCE/DOCKET NUMBER: BNL93-28A
;; TELEPHONE: (516) 282-7338
;; TELEFAX: (516) 282-3729
;; INFORMATION FOR SEQ ID NO: 52:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 336 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-235-836C-52

Query Match 1.2%; Score 7; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
Db 163 SLGSGQA 169

RESULT 13
US-08-235-836C-54
; Sequence 54, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235.836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93

;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bogosian, Margaret C.
;; REGISTRATION NUMBER: 25,324
;; REFERENCE/DOCKET NUMBER: BNL93-28A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516) 282-7338
;; TELEFAX: (516) 282-3729
;; INFORMATION FOR SEQ ID NO: 54:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 336 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-235-836C-54

Query Match 1.2%; Score 7; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
Db 163 SLGSGQA 169

RESULT 14
US-09-196-293-14
; Sequence 14, Application US/09196293
; Patent No. 6183755
; GENERAL INFORMATION:
; APPLICANT: Fuchs, Renate
; APPLICANT: Motz, Manfred
; APPLICANT: Soutscheck, Erwin
; APPLICANT: Wilske, Bettina
; APPLICANT: Preac-Mursic, Vera
; TITLE OF INVENTION: Active proteins from Borrelia
; FILE OF INVENTION: burgdorferi
; FILE REFERENCE: 738.001US2
; CURRENT APPLICATION NUMBER: US/09/196,293
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: US 08/209,603
; EARLIER FILING DATE: 1994-03-10
; EARLIER APPLICATION NUMBER: US 07/862,535
; EARLIER FILING DATE: 1992-06-19
; EARLIER APPLICATION NUMBER: WO PCT/EP90/02282
; EARLIER FILING DATE: 1990-12-21
; EARLIER APPLICATION NUMBER: DE P39 42 728.5
; EARLIER FILING DATE: 1989-12-22
; EARLIER APPLICATION NUMBER: DE P40 18 988.0
; EARLIER FILING DATE: 1990-06-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-09-196-293-14

Query Match 1.2%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSGQA 546
Db 168 SLGSGQA 174

RESULT 15
US-08-209-603E-14
; Sequence 14, Application US/08209603E
; Patent No. 6248538
; GENERAL INFORMATION:

APPLICANT: FUCHS, RENATE
APPLICANT: WILKE, BETTINA
APPLICANT: PREAC-MURSK, VERA
APPLICANT: MOTZ, MANFRED
APPLICANT: SOUTSCHECK, ERWIN
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PROTEINS
TITLE OF INVENTION: FROM BORRELIA BURGDORFERI
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROOKS HALDT HAFNER & DELAHUNTY
STREET: 99 PARK AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISC
COMPUTER: AT&T - IBM COMPATIBLE
OPERATING SYSTEM: MS-DOS Version 6.2
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,603E
FILING DATE: 10-MAR-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP90/02282
FILING DATE: 21-DEC-1990
APPLICATION NUMBER: US 07/862,535
FILING DATE: 19-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: ROBINSON, WILLIAM R.
REGISTRATION NUMBER: 27,224
REFERENCE/DOCKET NUMBER: LAR-9217-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-3355
TELEFAX: (212) 557-3635
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 341
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
DESCRIPTION: N/A
HYPOTHETICAL: N/A
ANTI-SENSE: N/A
FRAGMENT TYPE: N/A
ORIGINAL SOURCE:
ORGANISM: B. BURGDORFERI
IMMEDIATE SOURCE:
LIBRARY: DSM 5662
POSITION IN GENOME: N/A
FEATURE:
IDENTIFICATION METHOD: amino acid analysis
PUBLICATION INFORMATION: N/A
US-08-209-603E-14

Query Match 1.2%; Score 7; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLGSQA 546
Db 168 SLGSQA 174

RESULT 16
US-08-500-222-6
Sequence 6, Application US/08500222
Patent No. 5643733
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J

APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
TITLE OF INVENTION: and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704
FILING DATE: 21-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31
US-08-500-222-6
Query Match 1.2%; Score 7; DB 1; Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLGSQA 546
Db 277 SLGSQA 283
RESULT 17
US-08-500-125-6
Sequence 6, Application US/08500125
Patent No. 5643751
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
TITLE OF INVENTION: and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,704
; FILING DATE: 21-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: B31
; US-08-500-125-6

```

```

Query Match      1.2%  Score 7:  DB 1;  Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 540 SLSGSQA 546
   |||||
Db 277 SLSGSQA 283

```

```

RESULT 18
US-07-779-704B-6
; Sequence 6, Application US/07779704B
; Patent No. 5965702
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tam J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/779,704B
; FILING DATE: 19911021
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: B31
; US-07-779-704B-6

```

```

Query Match      1.2%  Score 7:  DB 2;  Length 376;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 540 SLSGSQA 546
   |||||
Db 277 SLSGSQA 283

```

```

RESULT 19
US-07-688-352C-16
; Sequence 16, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3836
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 464 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-688-352C-16

```

Query Match 1.2%; Score 7; DB 1; Length 464;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
 Db 440 FVYKRIK 446

RESULT 20
 US-08-474-379C-16
 ; Sequence 16, Application US/08474379C
 ; Patent No. 5977305
 ; GENERAL INFORMATION:
 ; APPLICANT: Wigler, Michael H.
 ; APPLICANT: Colicelli, John J.
 ; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
 ; TITLE OF INVENTION: PROCESSES
 ; NUMBER OF SEQUENCES: 88
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive/6300 Sears Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/474,379C
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/511,715
 ; FILING DATE: 20-APR-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/206,188
 ; FILING DATE: 01-MAR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/688,352
 ; FILING DATE: 19-APR-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clough, David W.
 ; REGISTRATION NUMBER: 36,107
 ; REFERENCE/DOCKET NUMBER: 27866/32771
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 474-6300
 ; TELEFAX: (312) 474-0448
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 464 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-474-379C-16

Query Match 1.2%; Score 7; DB 2; Length 464;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
 Db 440 FVYKRIK 446

RESULT 21
 US-09-146-249A-16

; Sequence 16, Application US/09146249A
 ; Patent No. 6069240
 ; GENERAL INFORMATION:
 ; APPLICANT: Wigler, Michael H.
 ; APPLICANT: Colicelli, John J.
 ; TITLE OF INVENTION: Cloning by Complementation and Related
 ; TITLE OF INVENTION: Processes
 ; NUMBER OF SEQUENCES: 85
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/146,249A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/511,715
 ; FILING DATE: 20-APR-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clough, David W.
 ; REGISTRATION NUMBER: 36,107
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312-474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 464 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-146-249A-16

Query Match 1.2%; Score 7; DB 3; Length 464;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
 Db 440 FVYKRIK 446

RESULT 22
 US-08-206-168B-16
 ; Sequence 16, Application US/08206188B
 ; Patent No. 6100025
 ; GENERAL INFORMATION:
 ; APPLICANT: Wigler, Michael H.
 ; APPLICANT: Colicelli, John J.
 ; TITLE OF INVENTION: Cloning by Complementation and Related
 ; TITLE OF INVENTION: Processes
 ; NUMBER OF SEQUENCES: 84
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206.188B
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511.715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-206-188B-16

Query Match 1.2%; Score 7; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
DB 440 FVYKRIK 446

RESULT 23

PCT-US91-02714-16
Sequence 16, Application PC/TUS9102714
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 464 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-02714-16

Query Match 1.2%; Score 7; DB 5; Length 464;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 FVYKRIK 483
DB 440 FVYKRIK 446

RESULT 24

US-08-500-222-2
Sequence 2, Application US/08500222
Patent No. 5643733
GENERAL INFORMATION:
APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
TITLE OF INVENTION: and uses thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/500,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704
FILING DATE: 21-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: B31
US-08-500-222-2

Query Match 1.2%; Score 7; DB 1; Length 498;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSSGSA 546

Db 350 SLGSQA 356
|||||||

RESULT 25

US-08-500-125-2
; Sequence 2, Application US/08500125
; Patent No. 5643751

GENERAL INFORMATION:

APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C
TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/500,125
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/779,704

FILING DATE: 21-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 498 amino acids
TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Borrelia burgdorferi
STRAIN: B31

US-08-500-125-2

Query Match

Best Local Similarity 1.2%; Score 7; DB 1; Length 498;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLGSQA 546

Db 350 SLGSQA 356

|||||||

RESULT 26

US-07-779-704B-2

; Sequence 2, Application US/07779704B
; Patent No. 5965702

GENERAL INFORMATION:

APPLICANT: Robinson, John M
APPLICANT: Pilot-Matias, Tami J
APPLICANT: Hunt, Jeffrey C

; TITLE OF INVENTION: Borrelia burgdorferi antigens
and uses thereof
NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories
STREET: One Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/779,704B
FILING DATE: 19911021

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Wong, Wean Khing
REGISTRATION NUMBER: 33561
REFERENCE/DOCKET NUMBER: 5051.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396
TELEFAX: 708-937-9556

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 498 amino acids
TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

ORGANISM: Borrelia burgdorferi
STRAIN: B31

US-07-779-704B-2

Query Match

Best Local Similarity 1.2%; Score 7; DB 2; Length 498;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 SLGSQA 546

Db 350 SLGSQA 356

|||||||

RESULT 27

US-08-796-899-25

; Sequence 25, Application US/08796899
; Patent No. 6160202

GENERAL INFORMATION:

APPLICANT: BUSTOS, Mauricio M

APPLICANT: CHERN, Maw-Sheng

TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH

TITLE OF INVENTION: TRANSCRIPTION FACTORS

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/796,899
FILING DATE: 06-FEB-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/319,544
FILING DATE: 07-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 028754-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 514 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-796-899-25

Query Match 1.2%; Score 7; DB 4; Length 514;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 SIKKELE 153
|||||
Db 483 SIKKELE 489

RESULT 28
US-09-188-579-62
Sequence 62, Application US/09188579B
Patent No. 6107040
GENERAL INFORMATION:
APPLICANT: Shuman, Stewart
TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
FILE REFERENCE: D6185
CURRENT APPLICATION NUMBER: US/09/188,579B
CURRENT FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 114
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: vaccinia virus
FEATURE:
OTHER INFORMATION: Motif III of RNA guanylyltransferase.
US-09-188-579-62

Query Match 1.0%; Score 6; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 VVVFGE 277
|||||
Db 1 VVVFGE 6

RESULT 29
US-08-960-780-69
Sequence 69, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Feitelson, Jerald S.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeits, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph

APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-960-780-69

Query Match 1.0%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 398 KGRVAV 403
|||||
Db 2 KGRVAV 7

RESULT 30
US-09-315-444-62
Sequence 62, Application US/09315444A
Patent No. 6232070
GENERAL INFORMATION:
APPLICANT: Shuman, Stewart
TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
FILE REFERENCE: D6185CIP
CURRENT APPLICATION NUMBER: US/09/315,444A
CURRENT FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: US 09/188,579
PRIOR FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: vaccinia virus
FEATURE:
OTHER INFORMATION: Motif III of RNA guanylyltransferase.
US-09-315-444-62

Query Match 1.0%; Score 6; DB 4; Length 8;

Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 VVVEGE 277
Db 1 VVVEGE 6

RESULT 31
US-09-073-898-69
; Sequence 69, Application US/09073898
; Patent No. 6242669
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeits, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-Cohn, Judy
; APPLICANT: Stamp, Lisa
; APPLICANT: Morrill, George
; APPLICANT: Finstad-Lee, Stacey
; TITLE OF INVENTION: No. 6242669el Pesticidal Toxins and Nucleotide
; TITLE OF INVENTION: Sequences Which Encode These Toxins
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-073-898-69

Query Match 1.0%; Score 6; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 398 KGRVAV 403
Db 2 KGRVAV 7

RESULT 32
US-08-159-339A-1037
; Sequence 1037, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-0050300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 1037:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-1037

Query Match 1.0%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 LVKEKG 160
Db 4 LVKEKG 9

RESULT 33
US-08-480-190-162
; Sequence 162, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley

APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,190
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,255
FILING DATE: June 15, 1993
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-480-190-162

Query Match 1.0%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 34
US-08-488-379-162
Sequence 162, Application US/08488379
Patent No. 5860103
GENERAL INFORMATION:
APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chicz
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,379
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,255
FILING DATE: June 15, 1993
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-488-379-162

Query Match 1.0%; Score 6; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 35
PCT-US93-07545-162
Sequence 162, Application PC/TUS9307545
GENERAL INFORMATION:
APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chicz
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07545
FILING DATE: 19930811
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 162:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 15
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 PCT-US93-07545-162

Query Match 1.0%; Score 6; DB 5; Length 15;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
 Db 4 VYRAVT 9

RESULT 36
 US-08-480-190-161
 ; Sequence 161, Application US/08480190
 ; Patent No. 5827516
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert G. Urban
 ; APPLICANT: Roman M. Chicz
 ; APPLICANT: Dario A. A. Vignali
 ; APPLICANT: Mary L. Hedley
 ; APPLICANT: Lawrence J. Stern
 ; APPLICANT: Jack L. Strominger
 ; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
 ; NUMBER OF SEQUENCES: 274
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM PS/2 Model 502 or 55SX
 ; OPERATING SYSTEM: MS-DOS (Version 5.0)
 ; SOFTWARE: WordPerfect (Version 5.1)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/480,190
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/077,255
 ; FILING DATE: June 15, 1993
 ; APPLICATION NUMBER: 07/925,460
 ; FILING DATE: August 11, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Paul T.
 ; REGISTRATION NUMBER: 30,162
 ; REFERENCE/DOCKET NUMBER: 00246/168001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-8906
 ; TELEFAX: (617) 542-5070
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 161:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 17
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-480-190-161

QY 236 VYRAVT 241
 Db 4 VYRAVT 9

RESULT 37
 US-08-488-379-161
 ; Sequence 161, Application US/08488379
 ; Patent No. 5880103
 ; GENERAL INFORMATION:
 ; APPLICANT: Robert G. Urban
 ; APPLICANT: Roman M. Chicz
 ; APPLICANT: Dario A. A. Vignali
 ; APPLICANT: Mary L. Hedley
 ; APPLICANT: Lawrence J. Stern
 ; APPLICANT: Jack L. Strominger
 ; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
 ; NUMBER OF SEQUENCES: 274
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM PS/2 Model 502 or 55SX
 ; OPERATING SYSTEM: MS-DOS (Version 5.0)
 ; SOFTWARE: WordPerfect (Version 5.1)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/488,379
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/077,255
 ; FILING DATE: June 15, 1993
 ; APPLICATION NUMBER: 07/925,460
 ; FILING DATE: August 11, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Paul T.
 ; REGISTRATION NUMBER: 30,162
 ; REFERENCE/DOCKET NUMBER: 00246/168001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-5070
 ; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 161:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 17
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-488-379-161

Query Match 1.0%; Score 6; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
 Db 4 VYRAVT 9

RESULT 38
 PCT-US93-07545-161
 ; Sequence 161, Application PC/TUS9307545
 ; GENERAL INFORMATION:

APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chic
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07545
FILING DATE: 19930811
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 161:
SEQUENCE CHARACTERISTICS:
LENGTH: 17
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07545-161

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 4 VYRAVT 9

RESULT 39
PCT-US95-08596-21
Sequence 21, Application PC/TUS9508596
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting
and Treating Type I Diabetes
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08596
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-092PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US95-08596-21

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative .0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEKG 160
Db 8 LVKEKG 13

RESULT 40
PCT-US95-08621-20
Sequence 20, Application PC/TUS9508621
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Invariant Chain Compositions Useful for
Treating Autoimmune Diseases
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08621
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-070PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
PCT-US95-08621-20

Query Match 1.0%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 LVKEKG 160
|||||
Db 8 LVKEKG 13

RESULT 41
US-08-480-190-160
; Sequence 160, Application US/08480190
; Patent No. 5827516
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,190
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-480-190-160

Query Match 1.0%; Score 6; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 VYRAVT 241
|||||
Db 5 VYRAVT 10

RESULT 42
US-08-488-379-160
; Sequence 160, Application US/08488379
; Patent No. 5880103
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
; NUMBER OF SEQUENCES: 274
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,379
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/077,255
; FILING DATE: June 15, 1993
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-488-379-160

Query Match 1.0%; Score 6; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 VYRAVT 241
|||||
Db 5 VYRAVT 10

RESULT 43
PCT-US93-07545-160
; Sequence 160, Application PC/TUS9307545
; GENERAL INFORMATION:
; APPLICANT: Robert G. Urban
; APPLICANT: Roman M. Chicz
; APPLICANT: Dario A. A. Vignali
; APPLICANT: Mary L. Hedley
; APPLICANT: Lawrence J. Stern
; APPLICANT: Jack L. Strominger
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES

; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07545
; FILING DATE: 19930811
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/925,460
; FILING DATE: August 11, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00246/168001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 160:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US93-07545-160

Query Match 1.0%; Score 6; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 VYRAVT 241
Db 5 VYRAVT 10

RESULT 44
US-08-468-543-10
; Sequence 10, Application US/08468543
; Patent No. 5726153
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,543
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,569
; FILING DATE: 16-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,929
; FILING DATE: 02-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,215
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,142
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/189,130
; FILING DATE: 02-MAY-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04547/002003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-468-543-10

Query Match 1.0%; Score 6; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 RELEAL 155
Db 12 RELEAL 17

RESULT 45
US-08-469-692-10
; Sequence 10, Application US/08469692
; Patent No. 5955055
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,692
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057

;; FILING DATE: 24-FEB-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/048,569
;; FILING DATE: 16-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/694,929
;; FILING DATE: 02-MAY-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/518,215
;; FILING DATE: 03-MAY-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/518,142
;; FILING DATE: 03-MAY-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/189,130
;; FILING DATE: 02-MAY-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Clark, Paul T.
;; REGISTRATION NUMBER: 30,162
;; REFERENCE/DOCKET NUMBER: 04547/002002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
US-08-469-692-10

Query Match 1.0%; Score 6; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 150 EELEAL 155
Db 12 BELEAL 17

Search completed: July 30, 2001, 11:43:18
Job time: 137 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:41:26 ; Search time 15.34 seconds
(without alignments)
2840.407 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFGKKSIPTSDRLIR.....RTAQKIWPPGGRNITLSLS 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	5.4	572	2 J49985	dihydropyrimidinase
2	31	5.4	572	2 JC5317	dihydropyrimidinase
3	31	5.4	572	2 S58889	collapsin response
4	18	3.1	570	2 JC5318	dihydropyrimidinase
5	18	3.1	570	2 S55525	Ulip protein - mou
6	18	3.1	572	2 JC5316	dihydropyrimidinase
7	14	2.4	519	2 JC5315	dihydropyrimidinase
8	14	2.4	519	2 S70581	dihydropyrimidinase
9	11	1.9	854	2 S33558	unc-33 protein - C
10	8	1.4	309	1 B64041	hypothetical prote
11	8	1.4	376	2 T51673	myb-related transc
12	8	1.4	465	2 A65071	hypothetical prote
13	8	1.4	465	2 F85942	hypothetical prote
14	8	1.4	520	2 T20007	hypothetical prote
15	8	1.4	546	2 A32260	cholesterol oxidas
16	8	1.4	648	2 T47988	serine/threonine-p
17	8	1.4	900	2 E83189	protein-PII uridyl
18	7	1.2	10	2 G60787	sperm-activating p
19	7	1.2	10	2 E60787	sperm-activating p
20	7	1.2	76	2 T51499	hypothetical prote
21	7	1.2	89	2 J00387	hypothetical prote
22	7	1.2	94	2 C69484	conserved hypothet
23	7	1.2	117	1 GAR82K	Ig heavy chain V-A
24	7	1.2	119	1 G1HUDW	Ig heavy chain V-I
25	7	1.2	123	2 H83903	hypothetical prote
26	7	1.2	124	2 S73210	ribosomal protein
27	7	1.2	127	2 C81292	hypothetical prote
28	7	1.2	135	2 C75465	conserved hypothet
29	7	1.2	138	2 C70077	hypothetical prote

conserved hypothet
RCc2 protein - ric
RCg2 protein - ric
hypothetical prote
hypothetical prote
type 4 fibrillar bi
hypothetical prote
probable lipoprote
hypothetical prote
hypothetical prote
troponin I - fruit
troponin I - fruit
hypothetical prote
p-nitrophenyl phos
sex-determining pr
SOS response regul
hypothetical prote

141 2 D69857
146 2 S53010
146 2 S53011
147 2 T16440
147 2 T30616
147 2 T30616
185 2 S77594
191 2 S76398
197 2 F81934
205 2 T33826
208 2 A38594
208 2 A40547
210 2 E72385
212 2 H84244
227 2 S35564
234 2 T35123
235 2 F70405

ALIGNMENTS

RESULT 1
S49985
dihydropyrimidinase-related protein 2 [similarity] - rat
N:Alternate names: collapsin response mediator protein 2; TONAD-64
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 12-May-1995 #text_change 02-Sep-2000
C:Accession: A59280; S49985
R:Minturn, J.E.; Fryer, H.J.; Geschwind, D.H.; Hockfield, S.
J. Neurosci. 15, 6757-6766, 1995
A:Title: TONAD-64, a gene expressed early in neuronal differentiation in the rat, is r
A:Reference number: A59280; MUID:96033765
A:Accession: A59280
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-572 <MIN>
A:Cross-references: GB:246882; NID:G599965; PIDN:CAA86981.1; PID:G599966
A:Experimental source: strain Sprague-Dawley; clone TONAD64 (turned on after division,
A:Note: submitted to the EMBL Data Library, November 1994
A:Note: in Genbank entry RRTONAD64, release 113.0, the source is designated as Rattus
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.7e-23;
Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSP 306
|||||
DB 276 GEPITASLGTGSHYWSKNWAKAAAFVTSP 306
|||||

RESULT 2
JC5317
dihydropyrimidinase-related protein 2 - human
N:Alternate names: collapsin response mediator protein 2
C:Species: Homo sapiens (man)
C:Date: 01-May-1997 #sequence_revision 23-Aug-1997 #text_change 02-Sep-2000
C:Accession: JC5317; S58891; JC7150
R:HamaJima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
Gene 180, 157-163, 1996
A:Title: A novel gene family defined by human dihydropyrimidinase and three related p
A:Reference number: JC5315; MUID:97128821
A:Accession: JC5317
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-572 <HAM>
A:Cross-references: DDBJ:D78013; NID:G1330239; PIDN:BA11191.1; PID:G1330240
A:Experimental source: fetal brain
R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
Nature 376, 509-514, 1995
A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
A:Reference number: S58889; MUID:95364923

A:Accession: S58891
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-572 <GOS>
A:Cross-references: EMBL:U17279; NID:gl244399; PIDN:AAA93202.1; PID:gl244400
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1996
R:Kitamura, K.; Takayama, M.; Hamajima, N.; Nakanishi, M.; Sasaki, M.; Endo, Y.; Takemoto
DNA Res. 6, 291-297, 1999
A:Title: Characterization of the human dihydropyrimidinase-related protein 2 (DRP-2) gene
A:Reference number: JC7150; MUID:20039612
A:Accession: JC7150
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-572 <KIT>
A:Cross-references: DBJ:AB020764; PIDN:BAH8422.1
C:Comment: This protein is the homologue of chicken 62K collapsin response mediator protein
C:Genetics:
A:Gene: GDB:DPYSL2; DRP2; CRMP2
A:Cross-references: GDB:5496401
A:Map position: 8p22-8p21
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.7e-23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
|||||
DB 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306

RESULT 3
S58889
collapsin response mediator protein, 62K - chicken
C:Species: Gallus gallus (chicken)
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000
C:Accession: S58889
R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
Nature 376, 509-514, 1995
A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein rel
A:Reference number: S58889; MUID:95364923
A:Accession: S58889
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-572 <GOS>
A:Cross-references: EMBL:U17277; NID:g882146; PIDN:AAA93200.1; PID:g882147
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 5.4%; Score 31; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.7e-23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306
|||||
DB 276 GEPITASLGTGSHYWSKNWAKAAAFVTSPP 306

RESULT 4
JC5318
dihydropyrimidinase related protein 3 - human
C:Species: Homo sapiens (man)
C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 02-Sep-2000
C:Accession: JC5318
R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
Gene 180, 157-163, 1996
A:Title: A novel gene family defined by human dihydropyrimidinase and three related prot
A:Reference number: JC5315; MUID:97128821
A:Accession: JC5318
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-570 <HAM>

A:Cross-references: DBJ:D78014; NID:gl330241; PIDN:BAAL1192.1; PID:gl330242
A:Experimental source: fetal brain
C:Comment: This protein is the homologue of rat 64K turned on after division protein
C:Genetics:
A:Gene: GDB:DPYSL3; DRP-3
A:Cross-references: GDB:5885804
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 570;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306
|||||
DB 289 HYWSKNWAKAAAFVTSPP 306

RESULT 5
S5525
Ulup protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 02-Sep-2000
C:Accession: S5525
R:Byk, T.; Dobransky, T.; Cifuentes-Diaz, C.; Sobel, A.
submitted to the EMBL Data Library, June 1995
A:Description: Identification and molecular characterisation of Ulup, a putative mamm
A:Reference number: S5525
A:Accession: S5525
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-570 <BYK>
A:Cross-references: EMBL:X87817; NID:g861058; PIDN:CAA61082.1; PID:g861059
C:Genetics:
A:Gene: Ulup
C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 570;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 HYWSKNWAKAAAFVTSPP 306
|||||
DB 289 HYWSKNWAKAAAFVTSPP 306

RESULT 6
JC5316
dihydropyrimidinase related protein 1 - human
C:Species: Homo sapiens (man)
C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 01-Dec-2000
C:Accession: JC5316; S58890
R:Hamajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Sasaki, M.; Nonaka, M.
Gene 180, 157-163, 1996
A:Title: A novel gene family defined by human dihydropyrimidinase and three related p
A:Reference number: JC5315; MUID:97128821
A:Accession: JC5316
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-572 <HAM>
A:Cross-references: DBJ:D78012; NID:gl330237; PIDN:BAAL1190.1; PID:gl330238
A:Experimental source: fetal brain
R:Goshima, Y.; Nakamura, F.; Strittmatter, P.; Strittmatter, S.M.
Nature 376, 509-514, 1995
A:Title: Collapsin-induced growth cone collapse mediated by an intracellular protein
A:Reference number: S58889; MUID:95364923
A:Accession: S58890
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 64-503, 'H', 505-572 <GOS>
A:Cross-references: EMBL:U17278; NID:g882148; PIDN:AAA93201.1; PID:g882149
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1994

C:Comment: This protein is involved in nervous system development.

C:Gene: GDB:CRMP1; DRP-1; DPYSL1

A:Cross-references: GDB:5585714

C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match 3.1%; Score 18; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 HYWKNWAKAAAFVTSPP 306

Db 289 HYWKNWAKAAAFVTSPP 306

RESULT 7

JC5315

diHydroPyrimidinase (EC 3.5.2.2) - human

N:Alternate names: 5,6-dihydropyrimidine amidohydrolase; Hydantoinase

C:Species: Homo sapiens (man)

C:Date: 01-May-1997 #sequence_revision 18-Jul-1997 #text_change 18-Aug-2000

C:Accession: JC5315

R:Hamaajima, N.; Matsuda, K.; Sakata, S.; Tamaki, N.; Nonaka, M.

A:Title: A novel gene family defined by human dihydropyrimidinase and three related proteins

A:Reference number: JC5315; MUID:97128821

A:Accession: JC5315

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-519 <HAM>

A:Cross-references: DDBJ:D78011; NID:g2339965; PIDN:BAAL1189.1; PID:g1330236

A:Experimental source: liver

C:Comment: This enzyme is Zn2+-metalloenzyme, and the second enzyme involved in uracil and dihydrothymine to N-carbamyl-beta-aminoisobutyrate. It catalyzes the hydrolysis of a variety of

C:Genetics:

A:Gene: GDB:DPYS; DHPase

A:Cross-references: GDB:5885803

C:Complex: homotetramer

C:Superfamily: allantoinase; Bacillus dihydroorotase homology

C:Keywords: hydrolase

Query Match

Best Local Similarity 100.0%; Pred. No. 9.6e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 380 FVAVTSTNAKIFN 393

Db 374 FVAVTSTNAKIFN 387

RESULT 8

S70581

diHydroPyrimidinase - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000

C:Accession: S70581

R:Matsuda, K.; Sakata, S.; Kaneko, M.; Hamaajima, N.; Nonaka, M.; Sasaki, M.; Tamaki, N.

A:Title: Molecular cloning and sequencing of a cDNA encoding dihydropyrimidinase from the

A:Reference number: S70581; MUID:96283806

A:Accession: S70581

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-519 <MAT>

A:Cross-references: EMBL:D63704; NID:g1378018; PIDN:BAA09833.1; PID:d1010479; PID:g13780

C:Superfamily: allantoinase; Bacillus dihydroorotase homology

Query Match

Best Local Similarity 100.0%; Pred. No. 9.6e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 QGKKAALAGGTTMI 104

Db 85 QGKKAALAGGTTMI 98

RESULT 9

S33558

unc-33 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 07-Apr-1994 #sequence_revision 02-Jun-1994 #text_change 21-Jul-2000

C:Accession: S33558; S33559; S24643; S24644

R:Li, W.; Herman, R.K.; Shaw, J.E.

Genetics 132, 675-689, 1992

A:Title: Analysis of the Caenorhabditis elegans axonal guidance and outgrowth gene un

A:Reference number: S33558; MUID:93106371

A:Accession: S33558

A:Molecule type: DNA

A:Residues: 1-854 <LIW1>

A:Cross-references: EMBL:Z14148; NID:g6899; PIDN:CAA78520.1; PID:g6900

A:Accession: S33559

A:Molecule type: mRNA

A:Residues: 1-854 <LIW2>

A:Cross-references: EMBL:Z14146; NID:g6903; PID:g6904

C:Genetics:

A:Introns: 14/2; 68/1; 129/3; 151/3; 182/1; 255/3; 361/2; 660/1; 790/3

C:Keywords: alternative initiators

F:1-854/Product: unc-33 protein (long form) #status predicted <MAT1>

F:176-854/Product: unc-33 protein (intermediate form) #status predicted <MAT2>

F:332-854/Product: unc-33 protein (short form) #status predicted <MAT3>

Query Match

Best Local Similarity 100.0%; Pred. No. 0.018;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 380 FVAVTSTNAAK 390

Db 683 FVAVTSTNAAK 693

RESULT 10

B64041

hypothetical protein H11730 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000

C:Accession: B64041

R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630

A:Accession: B64041

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-309 <TIGR>

A:Cross-references: GB:U032845; GB:L42023; NID:g3212236; PIDN:AAC23376.1; PID:g1574586

C:Superfamily: hypothetical protein H11730

Query Match

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 190 DLGALAQV 197

Db 264 DLGALAQV 271

RESULT 11

T51673

myb-related transcription factor MYB68 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: T51673
R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.; J.
Paz-Ares, J.; Weishaar, B.
Plant J. 16, 263-276, 1998
A:Title: Towards functional characterisation of the members of the R2R3-MYB gene from A.
A:Reference number: 214349; MUID:9839469
A:Accession: T51673
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-376 <KRA>
A:Cross-references: EMBL:AF062901; PIDN:AAC83623.1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: MYB68
A:Map position: V
C:Superfamily: Arabidopsis myb-related protein M4; myb DNA-binding repeat homology
C:Keywords: transcription factor

Query Match 1.4%; Score 8; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 AALAGGTT 102
|||||||
Db 337 AALAGGTT 344

RESULT 12

A65071
hypothetical protein b2873 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 22-Jun-1999
C:Accession: A65071
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617

A:Accession: A65071
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-465 <BLAT>
A:Cross-references: GB:A6000370; GB:U00096; NID:g2367170; PIDN:AAC75911.1; PID:g1789237;
A:Experimental source: strain K-12, substrain M61655
C:Superfamily: allantoinease; Bacillus dihydroorotase homology
F:49-452/Domain: Bacillus dihydroorotase homology <DHOS>

Query Match 1.4%; Score 8; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVDVHT 74
|||||||
Db 57 PGGVDVHT 64

RESULT 13

F85942
hypothetical protein 24212 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: F85942
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85942

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <STO>
A:Cross-references: GB:A6005174; NID:g12517394; PIDN:AAG58002.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z4212
C:Superfamily: allantoinease; Bacillus dihydroorotase homology

Query Match 1.4%; Score 8; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVDVHT 74
|||||||
Db 57 PGGVDVHT 64

RESULT 14

T20007
hypothetical protein C47E12.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T20007
R:Coles, L.
submitted to the EMBL Data Library, January 1996
A:Reference number: 219210
A:Accession: T20007
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-520 <WIL>
A:Cross-references: EMBL:Z68882; PIDN:CAA93104.1; GSPDB:GN00022; CESP:C47E12.8
A:Experimental source: clone C47E12
C:Genetics:
A:Gene: CESP:C47E12.8

A:Map position: 4
A:Introns: 29/2; 89/3; 147/1; 232/3; 310/2; 414/3
C:Superfamily: allantoinease; Bacillus dihydroorotase homology

Query Match 1.4%; Score 8; DB 2; Length 520;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 216 GITGPEGH 223
|||||||
Db 203 GITGPEGH 210

RESULT 15

A32260
cholesterol oxidase (EC 1.1.3.6) precursor [validated] - Streptomyces sp.
C:Species: Streptomyces sp.
C>Date: 20-Oct-1989 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: A32260; S15810; PC2002
R:Ishizaki, T.; Hirayama, N.; Shinkawa, H.; Niml, O.; Murooka, Y.
J. Bacteriol. 171, 596-601, 1989
A:Title: Nucleotide sequence of the gene for cholesterol oxidase from a Streptomyces
A:Reference number: A32260; MUID:89123081
A:Accession: A32260

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 'V', 2-546 <ISH>

A:Cross-references: GB:M31939; GB:J03356; NID:g153210; PIDN:AAA26719.1; PID:g153212
A:Note: the authors translated the initiation codon GTG for residue 1 as Val
R:Horii, M.; Ishizaki, T.; Paik, S.Y.; Manome, T.; Murooka, Y.
J. Bacteriol. 172, 3644-3653, 1990
A:Title: An operon containing the genes for cholesterol oxidase and a cytochrome P-45
A:Reference number: S15809; MUID:90299781
A:Accession: S15810
A:Molecule type: DNA
A:Residues: 1-30 <HOR>

A:Cross-references: EMBL:M31939; GB:J03356; NID:g153210
 R:Purcell, J.P.; Greenplate, J.T.; Jennings, M.G.; Kyerse, J.C.; Pershing, J.C.; Sims, S.
 Biochem. Biophys. Res. Commun. 196, 1406-1413, 1993
 A:Title: Cholesterol oxidase: a potent insecticidal protein active against boll weevil
 A:Reference number: PC2002; MUID:94071904
 A:Accession: PC2002
 A:Molecule type: protein
 A:Residues: 'XXXTF', 48-54, 'XX', 57, 'X', 59-60 <PUR>
 C:Genetics:
 A:Gene: choA
 C:Keywords: oxidoreductase
 F:1-42/Domain: signal sequence #status predicted <SIG>
 F:43-546/Product: cholesterol oxidase #status experimental <MAT>

Query Match 1.4%; Score 8; DB 2; Length 546;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
 |||||
 Db 21 AALAGGTT 28

RESULT 16
 T47988
 serine/threonine-protein kinase-like protein - Arabidopsis thaliana
 N:Alternate names: protein F21F14.130
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47988
 R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
 submitted to the Protein Sequence Database, February 2000
 A:Reference number: Z24481
 A:Accession: T47988
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-648 <CHO>
 A:Cross-references: EMBL:AL138642
 A:Experimental source: cultivar Columbia; BAC clone F21F14
 C:Genetics:
 A:Map position: 3
 A:Introns: 77/3; 117/1; 136/3; 158/2; 189/3; 216/3; 254/1; 272/2; 483/3; 563/1; 586/3
 A:Note: F21F14.130

Query Match 1.4%; Score 8; DB 2; Length 648;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 VNPDPPTA 314
 |||||
 Db 515 VNPDPPTA 522

RESULT 17
 E83189
 protein-PII uridylyltransferase PA3658 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83189
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuen, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337
 A:Accession: E83189
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-900 <STO>
 A:Cross-references: GB:AE004785; GB:AE004091; NID:g9949809; PIDN:AAG07046.1; GSPDB:GN001
 A:Experimental source: strain PA01

C:Genetics:
 A:Gene: glnD; PA3658
 C:Superfamily: uridylyltransferase

Query Match 1.4%; Score 8; DB 2; Length 900;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LQVTGSAH 333
 |||||
 Db 171 LQVTGSAH 178

RESULT 18
 G60787
 sperm-activating peptide (Ser-3,5,7 speract) - sea urchin (Anthocidaris crassispina)
 C:Species: Anthocidaris crassispina
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Aug-2000
 C:Accession: G60787
 R:Suzuki, N.; Kajiuura, H.; Nomura, K.; Garbers, D.L.; Yoshino, K.; Kurita, M.; Tanaka
 Comp. Biochem. Physiol. B 89, 687-693, 1988
 A:Title: Some more speract derivatives associated with eggs of sea urchins, Pseudocen
 A:Reference number: A60787; MUID:88242184
 A:Accession: G60787
 A:Molecule type: protein
 A:Residues: 1-10 <SUZ>
 C:Comment: This oligopeptide from egg jelly is one of several from this species, all
 at shows some, but not absolute, species restriction.
 C:Superfamily: unassigned animal peptides

Query Match 1.2%; Score 7; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 GFSLSGS 544
 |||||
 Db 1 GFSLSGS 7

RESULT 19
 E60787
 sperm-activating peptide (Ser-3,5,7, Asp-10 speract) - sea urchin (Hemicentrotus pulc
 C:Species: Hemicentrotus pulcherrimus
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Aug-2000
 C:Accession: E60787
 R:Suzuki, N.; Kajiuura, H.; Nomura, K.; Garbers, D.L.; Yoshino, K.; Kurita, M.; Tanaka
 Comp. Biochem. Physiol. B 89, 687-693, 1988
 A:Title: Some more speract derivatives associated with eggs of sea urchins, Pseudocen
 A:Reference number: A60787; MUID:88242184
 A:Accession: E60787
 A:Molecule type: protein
 A:Residues: 1-10 <SUZ>
 C:Comment: This oligopeptide from egg jelly is one of several from this species, all
 at shows some, but not absolute, species restriction.
 C:Superfamily: unassigned animal peptides

Query Match 1.2%; Score 7; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 4.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 538 GFSLSGS 544
 |||||
 Db 1 GFSLSGS 7

RESULT 20
 T51499
 hypothetical protein F5E19.10 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000

C:Accession: T51499
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A:Reference number: 225394
C:Accession: T51499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-76 <SAT>
A:Cross-references: EMBL:AL391147
A:Experimental source: cultivar Columbia; BAC clone F5E19
C:Genetics:
A:Map position: 5
A:Introns: 14/1; 60/2
A:Note: F5E19_10

Query Match 1.2%; Score 7; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EVEAEAV 236
|||||
Db 62 EVEAEAV 68

RESULT 21

JU0387
hypothetical 9.3K protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-Nov-1999
C:Accession: JU0387
R:Kanai, A.; Kuzuhara, T.; Sekimizu, K.; Natori, S.
J. Biochem. 109, 674-677, 1991
A:Title: Heterogeneity and tissue-specific expression of eukaryotic transcription factor
A:Reference number: JX0167; MUID:92011448
A:Molecule type: mRNA
A:Residues: 1-89 <KAN>
A:Cross-references: GB:D00925; NID:g220590; PIDN:BAA00767.1; PID:dl001229; PID:g220591

Query Match 1.2%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 QGRVALE 455
|||||
Db 66 QGRVALE 72

RESULT 22

C69484
conserved hypothetical protein AF1876 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69484
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 384-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: C69484
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-94 <KLE>
A:Cross-references: GB:AE000973; GB:AE000782; NID:g2689296; PIDN:AAB89376.1; PID:g264866

Query Match 1.2%; Score 7; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 155 LVKEGV 161
|||||
Db 85 LVKEGV 91
RESULT 23
GARB2K
Ig heavy chain V-A2 region (K-25) - rabbit (tentative sequence)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 31-Mar-2000
C:Accession: A02104
R:Jaton, J.C.
Biochem. J. 147, 235-247, 1975
A:Title: Comparison of the amino acid sequences of the variable domains of two homoge
A:Reference number: A90281; MUID:76039436
A:Accession: A02104
A:Molecule type: protein
A:Residues: 1-117 <JAT>
C:Comment: This chain was obtained from antibody to type III pneumococci and was isol
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: blocked amino end; heterotetramer; immunoglobulin
F:14-93/Domain: immunoglobulin homology <IMM>
F:1/Modified site: blocked amino end (Gln) (probably pyrrolidone carboxylic acid) #st
F:21-91/Disulfide bonds: #status experimental

Query Match 1.2%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
|||||
Db 24 SGFSLSG 30

RESULT 24

GIHUDW
Ig heavy chain V-II region (Daw) - human
C:Species: Homo sapiens (man)
C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 02-Sep-1997
C:Accession: A02091
R:Press, E.M.; Hogg, N.M.
Biochem. J. 117, 641-660, 1970
A:Title: The amino acid sequences of the Fd fragments of two human gamma heavy chain
A:Reference number: A90250; MUID:70258837
A:Accession: A02091
A:Molecule type: protein
A:Residues: 1-119 <PRE>
C:Comment: This chain was obtained from IgG1 isolated from the serum of a patient wit
C:Genetics:
A:Gene: GDB:IGHV6
A:Cross-references: GDB:128528; OMIM:147070
A:Map position: 14q32.33-14q32.33
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin; pyroglutamic acid
F:15-99/Domain: immunoglobulin homology <IMM>
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:22-97/Disulfide bonds: #status predicted
F:35-101/Disulfide bonds: #status experimental

Query Match 1.2%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSG 543
|||||
Db 25 SGFSLSG 31

RESULT 25

H83903
 hypothetical protein BH2032 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: H83903
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: AB3650; MUID:20263314
 A:Accession: H83903
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-123 <STO>
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05751.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2032
 C:Superfamily: Bacillus subtilis hypothetical protein ydhg

Query Match 1.2%; Score 7; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 ADAIAQA 267
 |||||
 Db 75 ADAIAQA 81

RESULT 26
 S73210
 ribosomal protein S12, chloroplast - red alga (Porphyra purpurea) chloroplast
 C:Species: chloroplast Porphyra purpurea
 C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 21-Jan-2000
 C:Accession: S73210
 R:Reith, M.; Munholland, J.
 Plant Mol. Biol. Rep. 13, 333-335, 1995
 A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
 A:Reference number: S73108
 A:Accession: S73210
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-124 <REI>
 A:Cross-references: EMBL:U38804; NID:g1276652; PIDN:AA08175.1; PID:g1276755
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1995
 C:Genetics:
 A:Gene: rps12
 A:Genome: chloroplast
 C:Superfamily: Escherichia coli ribosomal protein S12
 C:Keywords: chloroplast; protein biosynthesis; ribosome
 F:89/Modified site: beta-methylthioaspartic acid (Asp) #status predicted

Query Match 1.2%; Score 7; DB 2; Length 124;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LIRGRI 24
 |||||
 Db 81 LIRGRI 87

RESULT 27
 C81292
 hypothetical protein Cj1460 [imported] - Campylobacter jejuni (strain NCTC 11168)
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: C81292
 R:Parkhill, J.; Wren,
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: AB1250; MUID:20150912

A:Accession: C81292
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <PAR>
 A:Cross-references: GB:AL139078; GB:AL111168; NID:g96968723; PIDN:CAB73883.1; PID:g696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj1460

Query Match 1.2%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 EQRLLLE 214
 |||||
 Db 85 EQRLLLE 91

RESULT 28
 C75465
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: C75465
 R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C75465
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <WHI>
 A:Cross-references: GB:AE001941; GB:AE000513; NID:g6458589; PIDN:AAF10458.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0836
 A:Map position: 1

Query Match 1.2%; Score 7; DB 2; Length 135;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 VLPGGVD 71
 |||||
 Db 101 VLPGGVD 107

RESULT 29
 C70077
 hypothetical protein yxiG N17J [imported] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
 C:Accession: C70077; T47104
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardino
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033

A:Accession: C70077
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-138 <KUN>
 A:Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15955.1; PID:g2636465
 A:Experimental source: strain 158
 A:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
 Microbiology 141, 337-343, 1995
 A:Title: Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *ysj* gene
 A:Reference number: Z24350; MUID:95219088
 A:Accession: T47104
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-138 <YOS>
 A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06659.1; PID:g603785
 A:Experimental source: strain BGS1A1
 C:Genetics:
 A:Gene: *ysj*; N17J
 C:Superfamily: *Bacillus subtilis* hypothetical protein *ysj*
 Query Match 1.2% Score 7; DB 2; Length 138;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 DQSFYA 33
 Db 126 DQSFYA 132
 RESULT 30
 D69857
 conserved hypothetical protein *ykla* - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: D69857
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, R.; Riegler, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Serrano, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69857
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-141 <KUN>
 A:Cross-references: GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB13171.1; PID:g2633668
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: *ykla*
 C:Superfamily: hypothetical protein *ykla*
 Query Match 1.2% Score 7; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 151 ELEALVK 157
 Db 111 ELEALVK 117
 RESULT 31

S53010
 RCC2 protein - rice
 C:Species: *Oryza sativa* (rice)
 C:Date: 14-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 18-Feb-2000
 A:Accession: S53010
 R:Xu, Y.; Buchholz, W.G.; DeRose, R.T.; Hall, T.C.
 Plant Mol. Biol. 27, 237-248, 1995
 A:Title: Characterization of a rice gene family encoding root-specific proteins.
 A:Reference number: S53010; MUID:95195153
 A:Accession: S53010
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-146 <XUY>
 C:Superfamily: hydroxyproline-rich glycoprotein
 Query Match 1.2% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 441 GAPAVVI 447
 Db 32 GAPAVVI 38
 RESULT 32
 S53011
 RCg2 protein - rice
 C:Species: *Oryza sativa* (rice)
 C:Date: 14-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
 A:Accession: S53011
 R:Xu, Y.; Buchholz, W.G.; DeRose, R.T.; Hall, T.C.
 Plant Mol. Biol. 27, 237-248, 1995
 A:Title: Characterization of a rice gene family encoding root-specific proteins.
 A:Reference number: S53010; MUID:95195153
 A:Accession: S53011
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <XUY>
 A:Cross-references: GB:L27209; NID:g786129; PIDN:AAA65512.1; PID:g786130
 C:Superfamily: hydroxyproline-rich glycoprotein
 Query Match 1.2% Score 7; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 441 GAPAVVI 447
 Db 32 GAPAVVI 38
 RESULT 33
 T16440
 hypothetical protein F53A9.9 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
 A:Accession: T16440
 R:Miller, N.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid F53A9.
 A:Reference number: Z18513
 A:Accession: T16440
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-147 <MIL>
 A:Cross-references: EMBL:U23523; NID:g746551; PID:g746560; PIDN:AAC46564.1; CESP:F53A
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F53A9.9
 C:Superfamily: homeotic protein sal

Query Match 1.2% Score 7; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGGVVDVH 73
Db 110 PGGVVDVH 116

RESULT 34

T30616 hypothetical protein 14R - Molluscum contagiosum virus 1

N:Alternate names: MC014R
C:Species: Molluscum contagiosum virus 1

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 16-Feb-2001

C:Accession: T30616

R:Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.

Science 273, 813-816, 1996

A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re

A:Reference number: 220876; MUID:96325459

A:Accession: T30616

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-147 <SEN>

A:Cross-references: EMBL:U60315; NID:gl491943; PIDN:AAC55142.1; PID:gl491957

C:Genetics:

A:Note: MC014R

C:Superfamily: Molluscum contagiosum virus 1 hypothetical protein 14R

Query Match 1.2% Score 7; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 560 APPGGRS 566
Db 120 APPGGRS 126

RESULT 35

S77594

type 4 fimbrial biogenesis protein PilV PA4551 [imported] - Pseudomonas aeruginosa (stra

C:Species: Pseudomonas aeruginosa

C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 06-Oct-2000

C:Accession: S77594; A83076

R:Alm, R.A.; Mattick, J.S.

Mol. Microbiol. 16, 485-496, 1995

A:Title: Identification of a gene, pilV, required for type 4 fimbrial biogenesis in Pseu

A:Reference number: S77594; MUID:96037793

A:Accession: S77594

A:Molecule type: DNA

A:Residues: 1-185 <ALM>

A:Cross-references: EMBL:L36117

A:Experimental source: strain PA01

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Kas, A.; Larbig, K.; Lim

: Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: A83076

A:Molecule type: DNA

A:Residues: 1-185 <STO>

A:Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AAG07939.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: pilV; PA4551

C:Function:

A:Description: involved in biogenesis of type 4 fimbriae

C:Keywords: fimbria; membrane protein

Query Match 1.2% Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 534 LHQSGFS 540
Db 10 LHQSGFS 16

RESULT 36

S76398

hypothetical protein - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S76398

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys

S:Reference number: S74322; MUID:97061201

A:Accession: S76398

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-191 <KAN>

A:Cross-references: EMBL:D90915; GB:AB001339; NID:gl653604; PIDN:BAAL8527.1; PID:d101

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 1.2% Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ITSDRL 18
Db 156 ITSDRL 162

RESULT 37

F81934

probable lipoprotein NMA0883 [imported] - Neisseria meningitidis (strain 22491 serogr

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: F81934

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: A81775; MUID:20222556

A:Accession: F81934

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-197 <PAR>

A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84163.1; PID:g737

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: NMA0883

C:Superfamily: Neisseria meningitidis probable lipoprotein NMA0883

Query Match 1.2% Score 7; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125 ERADSA 131
Db 110 ERADSA 116

RESULT 38

T33826

hypothetical protein F23C8.6 - Caenorhabditis elegans

C:Species: *Caenorhabditis elegans*
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33826
R:Miller, N.; Wamsley, P.
submitted to the EMBL Data Library, November 1998
A:Description: The sequence of *C. elegans* cosmid F23C8.
A:Reference number: 221419
A:Accession: T33826
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-205 <MWL>
A:Cross-references: EMBL:AF106583; PIDN:AAD03134.1; GSPDB:GN00019; CESP:F23C8.6
A:Experimental source: strain Bristol N2; clone F23C8
C:Genetics:
A:Gene: CESP:F23C8.6
A:Map position: 1
A:Introns: 12/1; 47/3

Query Match 1.2%; Score 7; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AOVHAEN 201
|||||
DB 54 AOVHAEN 60

RESULT 39
A38594
troponin I - fruit fly (*Drosophila melanogaster*) (clone E2)
C:Species: *Drosophila melanogaster*
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 13-Aug-1999
C:Accession: A38594
R:Barbas, J.A.; Galceran, J.; Krah-Jentgens, I.; de la Pompa, J.L.; Canal, I.; Pongs, O.
Genes Dev. 5, 132-140, 1991
A>Title: Troponin I is encoded in the haplolethal region of the Shaker gene complex of *D. melanogaster*
A:Reference number: A38594; MUID:91115093
A:Accession: A38594
A:Molecule type: DNA
A:Residues: 1-208 <BAR>
A:Cross-references: GB:X58188; NID:g2511643; PIDN:CAA41171.1; PID:g8738
C:Genetics:
A:Gene: FlyBase:wupA
A:Cross-references: FlyBase:FBgn0004028
A:Introns: 5/3; 123/3; 185/3
C:Superfamily: troponin I

Query Match 1.2%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EEVEAEA 235
|||||
DB 202 EEVEAEA 208

RESULT 40
A40547
troponin I - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A40547
R:Beall, C.J.; Fyrberg, E.
J. Cell Biol. 114, 941-951, 1991
A>Title: Muscle abnormalities in *Drosophila melanogaster* heldup mutants are caused by mutations in the troponin I gene
A:Reference number: A40547; MUID:91340840
A:Accession: A40547
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-208 <BEA>
A:Cross-references: GB:X59376; NID:g8057; PIDN:CAA42020.1; PID:g8058

C:Genetics:
A:Gene: FlyBase:wupA
A:Cross-references: FlyBase:FBgn0004028
C:Superfamily: troponin I

Query Match 1.2%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EEVEAEA 235
|||||
DB 202 EEVEAEA 208

RESULT 41
E72385
hypothetical protein TM0375 - *Thermotoga maritima* (strain MSB8)
C:Species: *Thermotoga maritima*
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: E72385
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: E72385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <ARN>
A:Cross-references: GB:AE001717; GB:AE000512; NID:g4980871; PIDN:RAD35462.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0375
C:Superfamily: *Thermotoga maritima* hypothetical protein TM0375

Query Match 1.2%; Score 7; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEGV 161
|||||
DB 128 LVKEGV 134

RESULT 42
H84244
p-nitrophenyl phosphatase [imported] - *Halobacterium* sp. NRC-1
C:Species: *Halobacterium* sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84244
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A>Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: H84244
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <STO>
A:Cross-references: GB:AE004437; NID:gl0580446; PIDN:AAG19324.1; GSPDB:GN00138
C:Genetics:
A:Gene: pho2

Query Match 1.2%; Score 7; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 510 AKPGSGA 516
| | | | |
Db 114 AKPGSGA 120

RESULT 43

S35564
sex-determining protein SRY - common marmoset
C:Species: Callithrix jacchus (common marmoset)
C:Date: 20-May-1994 #sequence_revision 26-May-1995 #text_change 08-Sep-1997
C:Accession: S35564
R:Whitfield, L.S.; Lovell-Badge, R.; Goodfellow, P.N.
Nature 364, 713-715, 1993
A:Title: Rapid sequence evolution of the mammalian sex-determining gene SRY.
A:Reference number: S35558; MUID:9336117
A:Accession: S35564
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <WHI>
C:Genetics:
A:Gene: SRY
A:Map position: Y
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: DNA binding
F:57-132/Domain: HMG box homology <HMG>

Query Match 1.2%; Score 7; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 SPPVNP 310
| | | | |
Db 187 SPPVNP 193

RESULT 44

T35123
SOS response regulator lexA - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-Jan-2000
C:Accession: T35123
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1998
A:Reference number: 221568
A:Accession: T35123
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-234 <SEE>
A:Cross-references: EMBL:AL022268; PIDN:CAA18339.1; GSPDB:GN00070; SCOEDB:SC4H2.24c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: lexA; SCOEDB:SC4H2.24c
C:Superfamily: lexA protein

Query Match 1.2%; Score 7; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 AENGDIV 205
| | | | |
Db 178 AENGDIV 184

RESULT 45

F70405
hypothetical protein aq.1223 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: F70405
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: F70405
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-235 <AQF>
A:Cross-references: GB:AE000729; NID:g2983659; PIDN:AAC07236.1; PID:g2983670; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_1223

Query Match 1.2%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 KEELEAL 155
| | | | |
Db 37 KEELEAL 43

Search completed: July 30, 2001, 11:43:47
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:42:26 ; Search time 11.32 seconds
(without alignments)
1730.931 Million cell updates/sec

Title: US-09-367-496-8

Perfect score: 572

Sequence: 1 MSFQKKSIPRTSDRLIR.....RTAQKIMAPPGGRSNTLS 572

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	450	78.7	572	1	DPY4_HUMAN
2	55	9.6	564	1	DPY4_RAT
3	50	8.7	572	1	DPY4_MOUSE
4	31	5.4	572	1	DPY2_BOVIN
5	31	5.4	572	1	DPY2_CHICK
6	31	5.4	572	1	DPY2_HUMAN
7	31	5.4	572	1	DPY2_MOUSE
8	31	5.4	572	1	DPY2_RAT
9	18	3.1	358	1	DPY3_RAT
10	18	3.1	570	1	DPY3_HUMAN
11	18	3.1	570	1	DPY3_MOUSE
12	18	3.1	571	1	DPY3_XENLA
13	18	3.1	572	1	DPY1_HUMAN
14	18	3.1	572	1	DPY1_MOUSE
15	18	3.1	572	1	DPY1_RAT
16	14	2.4	519	1	DPY5_HUMAN
17	14	2.4	519	1	DPY5_RAT
18	11	1.9	854	1	UN33_CAEEL
19	8	1.4	309	1	YBGK_HAETIN
20	8	1.4	457	1	HYDA_AGRRD
21	8	1.4	465	1	YGE2_ECOLI
22	8	1.4	520	1	DPY1_CAEEL
23	8	1.4	546	1	CHOD_STRSQ
24	8	1.4	713	1	PALY_DIGLA
25	8	1.4	837	1	ATSA4_HUMAN
26	8	1.4	900	1	GLND_PSEAE
27	7	1.2	117	1	HV2B_RABIT
28	7	1.2	119	1	HV2C_HUMAN
29	7	1.2	124	1	RL2_PORPU
30	7	1.2	128	1	HSLU_SALTY
31	7	1.2	138	1	XYIG_BACSU
32	7	1.2	141	1	YKLA_BACSU
33	7	1.2	147	1	YV59_CAEEL

ALIGNMENTS

```

RESULT 1
DPY4_HUMAN
ID DPY4_HUMAN STANDARD; PRT; 572 AA.
AC Q14531; O00240;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYSL4 OR ULIP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-553 FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk I., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB006713; BAA21886.1; -
CC EMBL; Y10976; CAA71872.1; -
CC HSSP; P18316; 1KRC.
CC InterPro; IPR002195; -
CC Pfam; PF00744; Dihydroorotase; 1.
CC CONFLICT 122 122 R -> Q (IN REF. 2).
SQ SEQUENCE 572 AA; 61905 MW; 3E72A33E3BED5BE9 CRC64;

```

Query Match 78.7%; Score 450; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 WRIRADSAACCDYSLHVDITRWHSIKEEALVKKGNSFLVFWAYKDRCCSDSQMY 182
|||||
Db 123 WRIRADSAACCDYSLHVDITRWHSIKEEALVKKGNSFLVFWAYKDRCCSDSQMY 182
|||||

QY 183 EITSIIRDLGALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRAVTI 242
|||||
Db 183 EITSIIRDLGALAQVHAENGDIIVEEQKRLLELIGITGPEGHVLSHPPEVEAEAVYRAVTI 242
|||||

068564 pseudomonas
P53633 saccharomyc
P51501 callithrix
P15951 homo sapien
P39577 bacillus su
P24144 rhizobium l
Q59112 acidaminoco
P71887 mycobacteri
P34516 caenorhabdi
P28163 strongyloce
Q26846 strongyloce
O07637 bacillus su

QY 243 AKQANCLYVTVKMSKGRADATAQAQRGVVVGEPITASLTGDSHYWKNWAKAAAFV 302
DB 243 AKQANCLYVTVKMSKGRADATAQAQRGVVVGEPITASLTGDSHYWKNWAKAAAFV 302
QY 303 TSPPNPDPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIERM 362
DB 303 TSPPNPDPTTADHLTCLSSGDLQVTSAGHCTFTTAQKAVGKDNFALIPETNGIERM 362
QY 363 SMWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNPKAUKIISA 422
DB 363 SMWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNPKAUKIISA 422
QY 423 KTHNLNVEYNIPEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDPFVKRI 482
DB 423 KTHNLNVEYNIPEGVECRGAPAVVISQGRVALEDGKMFVTPGAGRVPRKTPDPFVKRI 482
QY 483 KARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRNLHQSGFSL 542
DB 483 KARNRLAEITHGVPRGLYDGPVHEVMVPAKPGSGAPARASCPCGKISVPPVRNLHQSGFSL 542
QY 543 GSOADHIAARTQAQIMAPPGGRSNTSLS 572
DB 543 GSOADHIAARTQAQIMAPPGGRSNTSLS 572

RESULT 2
ID DPY4_RAT STANDARD; PRT; 564 AA.
AC Q62951;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 3) (CRMP-3) (FRAGMENT).
GN DPYSL4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8915901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
RT nervous system";
RL J. Neurosci. 16:6197-6207(1996).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED TRANSIENTLY IN DEVELOPING SPINAL
CC CORD AND SELECTIVELY IN THE POSTNATAL CEREBELLUM.
CC -!- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U52103; AAB03281.1; -;
DR HSSP; P18316; IKRC.
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT NON_TER 1
SQ SEQUENCE 564 AA; 61085 MW; 42050891CC1436D2 CRC64;

Query Match 9.6%; Score 55; DB 1; Length 564;
Best Local Similarity 100.0%; Pred. No. 4.1e-49;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TIDAHGLAVLPQGVVDVHTRLQMPVLGMPADDFCOGTRKAALAGGTTMILDRHVPD 111
DB 49 TIDAHGLAVLPQGVVDVHTRLQMPVLGMPADDFCOGTRKAALAGGTTMILDRHVPD 103
RESULT 3
ID DPY4_MOUSE STANDARD; PRT; 572 AA.
AC O35098; O08886;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYSL4 OR ULIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kowaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB006715; BAA21888.1; -;
DR EMBL; Y09079; CAA70299.1; -;
DR MGD; MGI:1349764; Dpysl4.
DR InterPro; IPR002195; -;
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 125 126 ER -> DG (IN REF. 2).
FT CONFLICT 354 354 G -> V (IN REF. 2).
FT CONFLICT 420 420 F -> I (IN REF. 2).
SQ SEQUENCE 572 AA; 61961 MW; 37671129FC02C7AF CRC64;

Query Match 8.7%; Score 50; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 7e-44;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 VWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNP 414
DB 365 VWEKCVASGKMDNEFVAVTSTNAKIFNFPYPRKGRVAVGSDADLVWNP 414

RESULT 4
ID DPY2_BOVIN STANDARD; PRT; 572 AA.
AC O02675;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (NEURAL SPECIFIC PROTEIN
DE NSP60).
GN DPYSL2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kanata T.K.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U83278; AAB80618.1; -
 DR HSP; P18316; IKRC.
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 572 AA; 62277 MW; 343507ACB9D91BDE CRC64;
 CC
 CC Query Match 5.4%; Score 31; DB 1; Length 572;
 CC Best Local Similarity 100.0%; Pred. No. 5e-24;
 CC Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 Qy 276 GEPITASLGTDGSHYKSNWAKAAAFVTSPP 306
 Db 276 GEPITASLGTDGSHYKSNWAKAAAFVTSPP 306
 CC
 CC RESULT 5
 CC ID DPY2_CHICK STANDARD; PRT; 572 AA.
 CC AC Q90635;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSIN RESPONSE
 CC MEDIATOR PROTEIN CRMP-62).
 CC OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCB1_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Dorsal root ganglion;
 RX MEDLINE=95364923; PubMed=7637782;
 RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
 RT "Collapsin-induced growth cone collapse mediated by an intracellular
 RL protein related to UNC-33.";
 RL Nature 376:509-514(1995).
 CC -1- FUNCTION: INVOLVED IN NEURONAL GROWTH CONE COLLAPSE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U17277; AAA93200.1; -
 DR HSP; P18316; IKRC.
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 572 AA; 62330 MW; 85DB9E3DD5E54D8D CRC64;

Query Match 5.4%; Score 31; DB 1; Length 572;
 Best Local Similarity 100.0%; Pred. No. 5e-24;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 Qy 276 GEPITASLGTDGSHYKSNWAKAAAFVTSPP 306
 Db 276 GEPITASLGTDGSHYKSNWAKAAAFVTSPP 306
 CC

CC RESULT 6
 CC ID DPY2_HUMAN STANDARD; PRT; 572 AA.
 CC AC Q16555; O00424;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 01-OCT-2000 (Rel. 40, Last annotation update)
 CC DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (COLLAPSIN RESPONSE
 CC MEDIATOR PROTEIN 2) (CRMP-2) (N2A3).
 CC GN DPYSL2 OR CRMP2.
 CC OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95364923; PubMed=7637782;
 RA Goshima Y., Nakamura F., Strittmatter P., Strittmatter S.M.;
 RT "Collapsin-induced growth cone collapse mediated by an intracellular
 RL protein related to UNC-33.";
 RL Nature 376:509-514(1995).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97128821; PubMed=8973361;
 RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
 RT "A novel gene family defined by human dihydropyrimidinase and three
 RL related proteins with differential tissue distribution.";
 RL Gene 180:157-163(1996).
 CC [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Zhou J., Chen Y., Gu J.R.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20039612; PubMed=10574455;
 RA Kitamura K., Takayama M., Hamajima N., Nakanishi M., Sasaki M.,
 RA Endo Y., Takemoto T., Kimura H., Iwaki M., Nonaka M.;
 RT "Characterization of the human dihydropyrimidinase-related protein 2
 RL (DRP-2) gene.";
 RL DNA Res. 6:291-297(1999).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U7279; AAA93202.1; -
 DR EMBL; D78013; BAA11191.1; -
 DR EMBL; U97105; AAC05793.1; -
 DR EMBL; AB020777; BAA86991.1; -
 DR EMBL; AB020764; BAA86991.1; JOINED.
 DR EMBL; AB020765; BAA86991.1; JOINED.
 DR EMBL; AB020766; BAA86991.1; JOINED.
 DR EMBL; AB020767; BAA86991.1; JOINED.

```
DR EMBL; AB020768; BAA86991.1; JOINED.
DR EMBL; AB020769; BAA86991.1; JOINED.
DR EMBL; AB020770; BAA86991.1; JOINED.
DR EMBL; AB020771; BAA86991.1; JOINED.
DR EMBL; AB020772; BAA86991.1; JOINED.
DR EMBL; AB020773; BAA86991.1; JOINED.
DR EMBL; AB020774; BAA86991.1; JOINED.
DR EMBL; AB020775; BAA86991.1; JOINED.
DR EMBL; AB020776; BAA86991.1; JOINED.
DR HSSP; P18316; 1KRC.
DR MIN; 602463; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62293 MW; 5CDB6CF7F5C308AD CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 7
DPY2_MOUSE
ID DPY2_MOUSE STANDARD; PRT; 572 AA.
AC Q08553;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (ULIP 2 PROTEIN).
GN DPYSL2 OR ULIP2 OR CRMP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y10339; CAA71370.1; -.
DR MGD; MGI:1349763; Dpysl2.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62170 MW; 7FD4E8A242ACF62D CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 8
DPY3_RAT
ID DPY3_RAT STANDARD; PRT; 358 AA.
AC Q62952;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 4) (CRMP-4) (FRAGMENT).
GN DPYSL3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
```

```
DPY2_RAT
ID DPY2_RAT STANDARD; PRT; 572 AA.
AC P47942;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-2 (DRP-2) (TURNED ON AFTER
DE DIVISION, 64 KDA PROTEIN) (TOAD-64) (COLLAPSPIN RESPONSE MEDIATOR
DE PROTEIN 2) (CRMP-2).
GN DPYSL2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 136-142; 402-418; 441-450 & 499-511.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96033765; PubMed=7472434;
RA Minturn J.E., Fryer H.J.L., Geschwind D.H., Hockfield S.;
RT "TOAD-64, a gene expressed early in neuronal differentiation in the
RT rat, is related to unc-33, a C. elegans gene involved in axon
RT outgrowth.";
RL J. Neurosci. 15:6757-6766(1995).
CC -1- FUNCTION: MAY HAVE A ROLE IN AXON ELABORATION.
CC -1- SUBCELLULAR LOCATION: TIGHTLY, BUT NONCOVALENTLY, ASSOCIATED WITH
CC MEMBRANES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY AFTER NEURONAL BIRTH
CC AND IS DRAMATICALLY DOWNREGULATED IN THE ADULT.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46882; CAA86981.1; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
SQ SEQUENCE 572 AA; 62277 MW; C031F3BC038AA737 CRC64;

Query Match          5.4%; Score 31; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 5e-24;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306
   |||||||
Db 276 GEPITASLTGDSHYWSKNWAKAAAFVTSPP 306

RESULT 9
DPY3_RAT
ID DPY3_RAT STANDARD; PRT; 358 AA.
AC Q62952;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-3 (DRP-3) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 4) (CRMP-4) (FRAGMENT).
GN DPYSL3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96424532; PubMed=8815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
```



```

DR EMBL; Y09080; CAA70300.1; -.
DR EMBL; AB006714; BAA21887.1; -.
DR MGD; MGI:107793; Crmp1.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
FT CONFLICT 338 338 T -> I (IN REF. 2).
FT CONFLICT 476 476 E -> K (IN REF. 3).
FT CONFLICT 489 489 F -> S (IN REF. 3).
FT CONFLICT 520 520 K -> E (IN REF. 2).
SQ SEQUENCE 572 AA; 62167 MW; FEI7DDCD735CAF8F CRC64;

Query Match          3.18; Score 18; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 1.9e-10; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

Qy 289 HYWSKNWAKAAAFVTSPP 306
|||||
Db 289 HYWSKNWAKAAAFVTSPP 306

RESULT 15
DPY1_RAT STANDARD; PRT; 572 AA.
AC Q62950; P70546;
DT 15-JUL-1998 (Rel. 36, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-1 (DRP-1) (COLLAPSPIN RESPONSE
DE MEDIATOR PROTEIN 1) (CRMP-1).
GN CRMP1 OR DPYSL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98424532; PubMed=9815901;
RA Wang L., Strittmatter S.M.;
RT "A family of rat CRMP genes is differentially expressed in the
RT nervous system.";
RL J. Neurosci. 16:6197-6207(1996).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Quach T.T., Honnorat J., Aguera M., Belin M.F., Kolattukudy P.E.,
RA Antoine J.C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U52102; AAB03280.1; -.
CC DR EMBL; U52095; AAB07042.1; -.
CC DR InterPro; IPR002195; -.
CC DR Pfam; PF00744; Dihydroorotase; 1. H -> Y (IN REF. 2).
FT CONFLICT 3 3 H -> Y (IN REF. 2).
SQ SEQUENCE 572 AA; 62195 MW; EDG3BD8C751CCDF CRC64;

Query Match          3.18; Score 18; DB 1; Length 572;
Best Local Similarity 100.0%; Pred. No. 1.9e-10; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

Qy 289 HYWSKNWAKAAAFVTSPP 306
|||||
Db 289 HYWSKNWAKAAAFVTSPP 306

RESULT 16
DPY5_HUMAN STANDARD; PRT; 519 AA.
AC Q14117;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DIHYDROPYRIMIDINASE (EC 3.5.2.2) (DHPASE) (HYDANTOINASE) (DHP).
GN DPY5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97128821; PubMed=8973361;
RA Hamajima N., Matsuda K., Sakata S., Tamaki N., Sasaki M., Nonaka M.;
RT "A novel gene family defined by human dihydropyrimidinase and three
RT related proteins with differential tissue distribution.";
RL Gene 180:157-163(1996).
RP [2]
RP SEQUENCE FROM N.A. AND VARIANTS DHP DEFICIENCY.
RX MEDLINE=98386086; PubMed=9718352;
RA Hamajima N., Kouwaki M., Vreken P., Matsuda K., Sumi S., Imaeda M.,
RA Ohba S., Kidouchi K., Nonaka M., Sasaki M., Tamaki N., Endo Y.,
RA de Abreu R., Rottevel J., van Kullenburg A., van Gennip A.,
RA Toqari H., Wada Y.;
RT "Dihydropyrimidinase deficiency: structural organization, chromosomal
RT localization, and mutation analysis of the human dihydropyrimidinase
RT gene.";
RL Am. J. Hum. Genet. 63:717-726(1998).
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
CC UREIDOPROPIONATE.
CC -1- TISSUE SPECIFICITY: LIVER AND KIDNEY.
CC -1- DISEASE: DEFECTS IN DPY5 ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
CC DISORDER CHARACTERIZED BY DIHYDROPYRIMIDINURIA AND ASSOCIATED WITH
CC A VARIABLE CLINICAL PHENOTYPE (EPILEPTIC OR CONVULSIVE ATTACKS,
CC DYSMORPHIC FEATURES AND SEVERE DEVELOPMENTAL DELAY, AND CONGENITAL
CC MICROVILLOUS ATROPHY).
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D78011; BAA1189.1; -.
CC DR EMBL; AB004678; BAA33067.1; JOINED.
CC DR EMBL; AB004669; BAA33067.1; JOINED.
CC DR EMBL; AB004670; BAA33067.1; JOINED.
CC DR EMBL; AB004671; BAA33067.1; JOINED.
CC DR EMBL; AB004672; BAA33067.1; JOINED.
CC DR EMBL; AB004673; BAA33067.1; JOINED.
CC DR EMBL; AB004674; BAA33067.1; JOINED.
CC DR EMBL; AB004675; BAA33067.1; JOINED.
CC DR EMBL; AB004676; BAA33067.1; JOINED.
CC DR EMBL; AB004677; BAA33067.1; JOINED.
CC DR MIN; 222748; -.
CC DR InterPro; IPR002195; -.
CC DR Pfam; PF00744; Dihydroorotase; 1.
CC DR Hydrolase; Disease mutation.
FT VARIANT 68 68 T -> R (IN DHP DEFICIENCY).
FT VARIANT 334 334 Q -> R (IN DHP DEFICIENCY).
FT VARIANT 68 68 T -> R (IN DHP DEFICIENCY).
FT VARIANT 334 334 Q -> R (IN DHP DEFICIENCY).
FT FT

```

FT VARIANT 360 360 W -> R (IN DHP DEFICIENCY).
 FT FTID=VAR_002269.
 FT VARIANT 435 435 G -> R (IN DHP DEFICIENCY).
 FT FTID=VAR_002270.
 FT VARIANT 490 490 R -> T (IN DHP DEFICIENCY).
 FT FTID=VAR_002271.
 SQ SEQUENCE 519 AA; 56629 MW; 882E33D7C49D6ECC CRC64;

Query Match 2.4%; Score 14; DB 1; Length 519;
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 FVAVTSTNAKIFN 393
 DB 374 FVAVTSTNAKIFN 387
 |||||

RESULT 17
 DPYS_RAT STANDARD; PRT; 519 AA.
 AC Q63150;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE DIHYDROPYRIMIDINASE (EC 3.5.2.2) (DHPASE) (HYDANTOINASE) (DHP).
 GN DPYS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RA MEDLINE=96283806; PubMed=8679696;
 RX Matsuda K., Sakata S., Kaneko M., Hamajima N., Nonaka M., Sasaki M.,
 RA Tanaki N.;
 RT "Molecular cloning and sequencing of a cDNA encoding
 RT dihydropyrimidinase from the rat liver."
 RL Biochim. Biophys. Acta 1307:140-144(1996).
 CC -!- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O = 3-
 CC UREIDOPROPIONATE.
 CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D63704; BAA09833.1; -
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Hydrolase.
 SQ SEQUENCE 519 AA; 56833 MW; CA4CEF46801FCF4B CRC64;

Query Match 2.4%; Score 14; DB 1; Length 519;
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 QGTRAAAGGTTMI 104
 DB 85 QGTRAAAGGTTMI 98
 |||||

RESULT 18
 UN33_CAEEL STANDARD; PRT; 854 AA.
 ID UN33_CAEEL
 AC Q01630;
 DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE UNC-33 PROTEIN.
 GN UNC-33.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=93106371; PubMed=1468626;
 RA Li W., Herman R.K., Shaw J.E.;
 RT "Analysis of the Caenorhabditis elegans axonal guidance and outgrowth
 RT gene unc-33."
 RL Genetics 132:675-689(1992).
 CC -!- FUNCTION: INVOLVED IN AXONAL GUIDANCE AND OUTGROWTH.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- ALTERNATIVE PRODUCTS: THREE PUTATIVE POLYPEPTIDES MAY ARISE BY
 CC ALTERNATIVE SPLICING.
 CC -!- DEVELOPMENTAL STAGE: UNC-33 PROTEINS ARE DISTRIBUTED EXCLUSIVELY
 CC WITHIN NEURONAL PROCESSES AFTER EARLY EMBRYOGENESIS.
 CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z14148; CAA78520.1; -
 DR EMBL; Z14148; CAA78521.1; -
 DR EMBL; Z14148; CAA78522.1; -
 DR EMBL; Z14146; CAA78516.1; -
 DR EMBL; Z14146; CAA78517.1; -
 DR EMBL; Z14146; CAA78518.1; -
 DR PIR; S24643; S24643.
 DR PIR; S24644; S24644.
 DR PIR; S33558; S33558.
 DR InterPro; IPR002195; -
 DR Pfam; PF00744; Dihydroorotase; 1.
 KW Alternative splicing.
 FT VARSPLIC 1 175 MISSING (IN 72.1 KDA ISOFORM).
 FT VARSPLIC 1 331 MISSING (IN 55.5 KDA ISOFORM).
 SQ SEQUENCE 854 AA; 90819 MW; A8073DDE251D2D77 CRC64;

Query Match 1.9%; Score 11; DB 1; Length 854;
 Best Local Similarity 100.0%; Pred. No. 0.0057;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 FVAVTSTNAK 390
 DB 683 FVAVTSTNAK 693
 |||||

RESULT 19
 YBCK_HAEIN STANDARD; PRT; 309 AA.
 ID YBCK_HAEIN
 AC P44298;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL PROTEIN H11730.
 GN H11730.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Buit C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frilichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RC -1- SIMILARITY: STRONG. TO E COLI YBCK.
CC -1- SIMILARITY: TO B.SUBTILIS YCSJ AND YEAST UREA AMIDOLYASE (DURI,2).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32845; AAC23376.1; -
DR TIGR; H11730; -
KW Hypothetical protein.
SQ SEQUENCE 309 AA; 34562 MW; E34D87B4C838EFID CRC64;

Query Match 1.4%; Score 8; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 DLGALAAQV 197
Db 264 DLGALAAQV 271

RESULT 20
HYDA_AGRD
ID HYDA_AGRD STANDARD; PRT; 457 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE D-HYDANTOINASE (EC 3.5.2.2) (DIHYDROPYRIMIDINASE) (DHPASE).
OS Agrobacterium radiobacter.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRLB 11291;
RA Grifantini R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 5,6-DIHYDROURACIL + H(2)O -> 3-
CC UREIDOPROPIONATE.
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91070; CAA62549.1; -
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
KW Hydrolase.

STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Buit C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frilichman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RC -1- SIMILARITY: STRONG. TO E COLI YBCK.
CC -1- SIMILARITY: TO B.SUBTILIS YCSJ AND YEAST UREA AMIDOLYASE (DURI,2).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32845; AAC23376.1; -
DR TIGR; H11730; -
KW Hypothetical protein.
SQ SEQUENCE 309 AA; 34562 MW; E34D87B4C838EFID CRC64;

Query Match 1.4%; Score 8; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 401 VAVGSDAD 408
Db 382 VAVGSDAD 389

RESULT 21
YGEZ_ECOLI
ID YGEZ_ECOLI STANDARD; PRT; 465 AA.
AC Q46806;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE HYPOTHETICAL 51.5 KDA PROTEIN IN PBL-LYSS INTERGENIC REGION.
GN YGEZ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997)
CC -1- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38375; AAC83054.1; -
DR EMBL; AB000370; AAC75911.1; -
DR EcoGene: EGI3056; YgeZ.
DR InterPro: IPR002195; -
DR Pfam: PF00744; Dihydroorotase; 1.
KW Hypothetical protein.
SQ SEQUENCE 465 AA; 51503 MW; F1B0C91922CE812D CRC64;

Query Match 1.4%; Score 8; DB 1; Length 465;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 PGQVDVHT 74
Db 57 PGQVDVHT 64

RESULT 22
DPY1_CAEEL
ID DPY1_CAEEL STANDARD; PRT; 520 AA.
AC Q18677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE HYPOTHETICAL 56.3 KDA PROTEIN C47E12.8 IN CHROMOSOME IV.
GN C47E12.8.
OS Caenorhabditis elegans.
```


DR EMBL; AJ002221; CAA05251.1; -
DR InterPro; IPR001106; -
DR Pfam; PF00221; PAL; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism; Multigene family.
FT SITE 198 200 MODIFIED TO FORM 4-METHYLLIDENE-IMIDAZOLE-
5-ONE (BY SIMILARITY).
SQ SEQUENCE 713 AA; 77732 MW; 747CB64C65BDC6A CRC64;

Query Match 1.4%; Score 8; DB 1; Length 713;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 RVALEDGK 458
Db 635 RVALEDGK 642

RESULT 25
ATSA_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE ADAM-TS 4 PRECURSOR (EC 3.4.24. -) (A DISINTEGRIN AND METALLOPROTEINASE
WITH THROMBOSPONDIN MOTIFS 4) (ADAMTS-4) (ADAM-TS4) (AGGREGANASE 1)
(ADMP-1).
DE ADAMTS4 OR KIAA0688.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
Liu R., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
Tzazakos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
family of proteases.";
RL Science 284:1664-1666(1999).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M., Pratta M., Liu R.O., Abbaszade I., Ross H., Burn T.,
Arner E.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
aggrecan substrate recognition and cleavage.";
RJ. Biol. Chem. 275:25791-25797(2000).
CC -|- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
CC -|- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 392-GLU-|-ALA-393
SITE.
CC -|- COPACITOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
MATRIX (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED

CC -|- INDUCTION: BY INTERLEUKIN 1.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014588; BAA31663.1; -
CC EMBL; AF148213; AAD41494.1; -
CC MIM; 603876; -
CC HSP; P34179; IIAG.
CC MEROPS; M12.221; -
CC InterPro; IPR000130; -
CC InterPro; IPR000884; -
CC InterPro; IPR001590; -
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS0092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837 ADAM-TS 4.
FT SITE 194 194 CYSTEINE SWITCH (POTENTIAL).
FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 362 362 BY SIMILARITY.
FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 437 519 DISINTEGRIN-LIKE.
FT DOMAIN 520 576 TSP-TYPE 1 1.
FT DOMAIN 577 685 CYS-RICH.
FT DOMAIN 686 837 SPACER.
FT DOMAIN 247 252 POLY-ALA.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 77 77 A -> T (IN REF. 1).
SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;

Query Match 1.4%; Score 8; DB 1; Length 837;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 PGSGAPAR 519
Db 73 PGSGAPAR 80

RESULT 26
GLND_PSEAE
ID GLND_PSEAE STANDARD; PRT; 900 AA.
AC Q929H0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE [PROTEIN-PII] URIDYLTRANSFERASE (EC 2.7.7.59) (PII URIDYL-
TRANSFERASE) (URIDYL REMOVING ENZYME) (UTASE).
GN GLND OR PA3658.
OS Pseudomonas aeruginosa.

CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PA01;
 RA Nashimoto H.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy R., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 FT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: MODIFIER, BY URIDYLATION OR DEURIDYLATION THE PII
 CC (GINS) REGULATORY PROTEIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UTP + [PROTEIN-PII] = PYROPHOSPHATE +
 CC URIDYL-[PROTEIN-PII].
 CC -1- SIMILARITY: BELONGS TO THE GINS FAMILY.
 CC -1- CAUTION: REF.1 DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN
 CC POSITION 285.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AB024601; BAA75913.1; ALT_FRAME.
 DR EMBL; AE004785; AAG07046.1; -
 DR InterPro; IPR002912; -
 DR InterPro; IPR002934; -
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF01909; NTP_transf_2; 1.
 DR Transferase; Nucleotidyltransferase.
 KW CONFLICT 811 811 L -> V (IN REF. 1).
 FT SEQUENCE 900 AA; 103404 MW; 7C71F31EC284E836 CRC64;
 SQ
 Query Match 1.4%; Score 8; DB 1; Length 900;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 326 LQVGTGSAH 333
 Db 171 LQVGTGSAH 178
 RESULT 27
 ID HV2B_RABIT STANDARD; PRT; 117 AA.
 AC P01828;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE IG HEAVY CHAIN V-A2 REGION K-25.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=76039436; PubMed=241319;
 RA Jaton J.-C.;

RT "Comparison of the amino acid sequences of the variable domains of
 RT two homogeneous rabbit antibodies to type III pneumococcal
 RT polysaccharide.";
 RL Biochem. J. 147:235-247(1975).
 CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM ANTIBODY TO TYPE III
 CC PNEUMOCOCCI AND WAS ISOLATED FROM THE SERUM OF A SINGLE RABBIT.
 DR PIR; A02104; GARB2K.
 DR InterPro; IPR003006; -
 DR Pfam; PF00047; ig; 1.
 KW Immunoglobulin V region.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT DISULFID 21 91
 FT NON_TER 117 117
 SQ SEQUENCE 117 AA; 12580 MW; 28DD87FDB7AEE9B8 CRC64;
 Query Match 1.2%; Score 7; DB 1; Length 117;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 537 SGFSLSG 543
 Db 24 SGFSLSG 30
 RESULT 28
 ID HV2C_HUMAN STANDARD; PRT; 119 AA.
 AC P01816;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE IG HEAVY CHAIN V-II REGION DAW.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=70258837; PubMed=5449120;
 RA Press E.M., Hogg N.M.;
 RT "The amino acid sequences of the Fd fragments of two human gamma-1
 RT heavy chains.";
 RL Biochem. J. 117:641-660(1970).
 CC -1- MISCELLANEOUS: THIS CHAIN WAS OBTAINED FROM IGG1 ISOLATED FROM THE
 CC SERUM OF A PATIENT WITH HYPERGAMMAGLOBULINEMIA.
 DR PIR; A02091; GIHUDW.
 DR InterPro; IPR003006; -
 DR Pfam; PF00047; ig; 1.
 KW Immunoglobulin V region.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT NON_TER 119 119
 SQ SEQUENCE 119 AA; 13045 MW; 4E13E00214BAD789 CRC64;
 Query Match 1.2%; Score 7; DB 1; Length 119;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 537 SGFSLSG 543
 Db 25 SGFSLSG 31
 RESULT 29
 ID RRI2_PORPU STANDARD; PRT; 124 AA.
 AC P51289;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S12.
 GN RPS12.

OS Porphyra purpurea.
 OG Chloroplast.
 CC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 CC NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AVONPORT;
 RA Reith M.E., Munnolland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 genome.";
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U38804; AAC08175.1; -;
 DR Mendel; 10344; PORPU; rps12; i.
 DR InterPro; IPR000230; -;
 DR Pfam; PF00164; Ribosomal_S12; 1.
 DR PRINTS; PRO1034; RIBOSOMALS12.
 DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
 KW Ribosomal protein; Chloroplast.
 SQ SEQUENCE 124 AA; 13904 MW; 6632BE623CAllFb1 CRC64;

 Query Match 1.2%; Score 7; DB 1; Length 124;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LIRGGRI 24
 D 18 LIRGGRI 87
 RESULT 30
 HSLU_SALTY STANDARD; PRT; 128 AA.
 ID HSLU_SALTY
 AC O30911;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ATP-DEPENDENT HSL PROTEASE ATP-BINDING SUBUNIT HSLU (HEAT SHOCK
 DE PROTEIN HSLU) (FRAGMENT).
 GN HSLU.
 OS Salmonella typhimurium.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97451029; PubMed=9302299;
 RA Valdivia R.H., Falkow S.;
 RT "Fluorescence-based isolation of bacterial genes expressed within
 RT host cells.";
 RL Science 277:2007-2011(1997).
 CC -1- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION
 CC COMPLEX (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH HSLV (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLXP CHAPERONE FAMILY. HSLU SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL: AF020812; AAB80745.1; -;
 DR StyGene; SG10655; hslu.
 KW Chaperone; ATP-binding; Heat shock.
 FT NON_TER 1
 FT NON_TER 128
 SQ SEQUENCE 128 AA; 14066 MW; 2378E6D9113BE41 CRC64;

 Query Match 1.2%; Score 7; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 557 KIMAPPG 563
 D 13 KIMAPPG 19
 RESULT 31
 YXIG_BACSU STANDARD; PRT; 138 AA.
 ID YXIG_BACSU
 AC P42299;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 16.2 KDA PROTEIN IN WAPA-LICT INTERGENIC REGION.
 GN YXIG OR N17J.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci.";
 RL Microbiology 141:337-343(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D51856; BAA06659.1; -;
 DR EMBL; D83026; BAA11686.1; -;
 DR EMBL; D29985; BAA06263.1; -;
 DR EMBL; Z59124; CAB15955.1; -;
 DR Subtilist; BG11136; yxig.
 KW Hypothetical protein.
 SQ SEQUENCE 138 AA; 16228 MW; DB806FB61EAF4D94 CRC64;

 Query Match 1.2%; Score 7; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 27 DDCSFYA 33
 D 126 DDCSFYA 132
 RESULT 32
 YKLA_BACSU STANDARD; PRT; 141 AA.
 ID YKLA_BACSU
 AC O34762;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 14.9 KDA PROTEIN IN PROA-METC INTERGENIC REGION.
GN YKLA.

OS Bacillus subtilis.

CC Bacteria; Firmicutes; Bacillus/clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Devine K.M.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE OSMC/OHR FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AJ002571; CAA05393.1; -

DR EMBL; Z99110; CAB13171.1; -

DR Subtilist; BG13238; YKLA.

KW Hypothetical protein.

SQ SEQUENCE 141 AA; 14868 MW; 7F4A32D0619D2233 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 141;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 ELEALVK 157

DB 111 ELEALVK 117

RESULT 33

ID YV59 CAEEL

AC P50439; STANDARD; PRT; 147 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE HYPOTHETICAL 15.2 KDA PROTEIN F53A9.9 IN CHROMOSOME X.

GN F53A9.9

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Miller N.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U23523; AAC46564.1; -

DR Wormpep; F53A9.9; CE02769.

KW Hypothetical protein.

SQ SEQUENCE 147 AA; 15184 MW; E08CBE80E91ADFCD CRC64;

Query Match

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 PGVDVH 73

DB 110 PGVDVH 116

RESULT 34

ID OMLA_PSEFL

AC O68564; STANDARD; PRT; 175 AA.

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE OUTER MEMBRANE LIPOPROTEIN OMLA PRECURSOR.

GN OMLA OR OPRX.

OS Pseudomonas fluorescens.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=294;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15453;

RA MEDLINE=99138728; PubMed-9973334;

RT "Pseudomonas aeruginosa fur overlaps with a gene encoding a novel

outer membrane lipoprotein, Omla.";

RL J. Bacteriol. 181:1099-1109(1999).

CC -1- FUNCTION: MAY HAVE A STRUCTURAL ROLE IN MAINTAINING THE CELL

CC ENVELOPE INTEGRITY.

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID

ANCHOR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SMPA/OMLA FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF050677; AAC05681.1; -

KW Outer membrane; Lipoprotein; Signal.

FT SIGNAL 1 21 BY SIMILARITY.

FT CHAIN 22 175 OUTER MEMBRANE LIPOPROTEIN OMLA.

FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).

FT DOMAIN 163 175 PRO-RICH.

SQ SEQUENCE 175 AA; 19123 MW; B03950C6931F57E7 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 175;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 VPAKPGS 514

DB 141 VPAKPGS 147

RESULT 35

ID YIPC_YEAST

AC P53633; STANDARD; PRT; 176 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE YIP3 PROTEIN.

GN YIP3 OR YNL044W OR N2650.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

```
RN RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE=96310629; PubMed=8740423;
RT "The sequence of 12.8 kb from the left arm of chromosome XIV reveals
RT a sigma element, a pro-tRNA and six complete open reading frames, one
RT of which encodes a protein similar to the human leukotriene A4
RT hydrolase."
RL Yeast 12:493-499(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Matern H.T., Gallwitz D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D., Hilbert H.,
RL Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97401; CAA66058.1; -
DR EMBL; X97400; CAA66057.1; -
DR EMBL; X94547; CAA64238.1; ALT_SEQ.
DR EMBL; Z71320; CAA95911.1; ALT_SEQ.
DR EMBL; Z71321; CAA95913.1; -
DR SGD; S0004989; YIP3.
KW Transmembrane.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 129 149 POTENTIAL.
SQ SEQUENCE 176 AA; 19446 MW; C88EA166EEDDE11B8 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LGALAQV 197
DB 4 LGALAQV 10

RESULT 36
SRY_CALJA
ID SRY_CALJA STANDARD; PRT; 227 AA.
AC P51501;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).
GN SRY.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
OC Callithrix.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93361117; PubMed=8355783;
RA Whitfield L.S., Lovell-Badge R., Goodfellow P.N.;
RL "Rapid sequence evolution of the mammalian sex-determining gene SRY."
RT Nature 364:713-715(1993).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC
CC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE
CC SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL
```

```
CC INTERCALATION IN THE MINOR GROOVE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X86386; CAA60146.1; -
DR HSSP; Q05086; 1HRZ.
DR InterPro; IPR000910; -
DR Pfam; PF00505; HMG_box; 1.
KW DNA-binding; Nuclear protein; Transcription regulation; Activator;
KW Sexual differentiation.
FT DNA_BIND 60 128 HMG_BOX.
SQ SEQUENCE 227 AA; 26307 MW; 94C5347D530C55AE CRC64;

Query Match 1.2%; Score 7; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 SPPVNPD 310
DB 187 SPPVNPD 193

RESULT 37
TRY3_HUMAN
ID TRY3_HUMAN STANDARD; PRT; 247 AA.
AC P15951;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRYPSIN III PRECURSOR (EC 3.4.21.4).
GN PRSS3 OR TRY3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90221895; PubMed=2326201;
RA Tani T., Kawashima I., Mita K., Takiguchi Y.;
RT "Nucleotide sequence of the human pancreatic trypsinogen III cDNA."
RL Nucleic Acids Res. 18:1631-1631(1990).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15505; CAA33527.1; -
DR PIR; S12764; S12764.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.174; -
DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
```

```

DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 23
FT CHAIN 24 247
FT ACT_SITE 63 63
FT ACT_SITE 107 107
FT ACT_SITE 200 200
FT ACT_SITE 30 160
FT DISULFID 48 64
FT DISULFID 139 206
FT DISULFID 171 185
FT SITE 194 194
SQ SEQUENCE 247 AA; 26776 MW; 697DE163F1CEE0D6 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 ASCPGKI 526
DB 173 ASCPGKI 179

RESULT 38
DLTE_BACSU STANDARD; PRT; 252 AA.
AC P39577;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DLTE PROTEIN.
GN DLTE OR IPA-1R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
RA Prescan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RA "Bacillus subtilis genome project: cloning and sequencing of the 97
FT kb region from 325 degrees to 333 degrees.";
RT Mol. Microbiol. 10:371-384(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95318144; PubMed=7797557;
RA Perego M., Glaser P., Minutello A., Strauch M.A., Leopold K.,
RA Fischer W.;
RT "Incorporation of D-alanine into lipoteichoic acid and wall teichoic
FT acid in Bacillus subtilis. Identification of genes and regulation.";
RL J. Biol. Chem. 270:15598-15606(1995).
CC -1- PATHWAY: BIOSYNTHESIS OF D-ALANYL-LIPOTEICHOIC ACID.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: TO 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASES.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73124; CAA51557.1;

```

```

DR EMBL; Z99123; CAB15880.1; -.
DR PIR; S39656; S39656.
DR Subtilist; BG10547; dlte.
DR InterPro; IPR002198; -.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase.
FT NP_BIND 9 33
FT ACT_SITE 152 152
FT ACT_SITE 152 152
SQ SEQUENCE 252 AA; 28268 MW; DDAF1BE457D5D62A CRC64;

Query Match 1.2%; Score 7; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 KRLLLEIG 216
DB 23 KRLLLEIG 29

RESULT 39
NODJ_RHILT STANDARD; PRT; 262 AA.
AC P24144;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NODULATION PROTEIN J.
GN NODJ.
OS Rhizobium leguminosarum (biovar trifolii).
OG Plasmid sym.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=386;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANU 843;
RX MEDLINE=90251164; PubMed=2338917;
RA Surin B.P., Watson J., Hamilton W.D.O., Economou A., Downie J.A.;
RT "Molecular characterization of the nodulation gene, nodT, from two
FT biovars of Rhizobium leguminosarum.";
RL Mol. Microbiol. 4:245-252(1990).
CC -1- FUNCTION: FORMS, WITH NODI, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ABC-2 SUBFAMILY OF INTEGRAL MEMBRANE
CC PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51411; CAA35772.1; -.
DR PIR; S08617; S08617.
DR InterPro; IPR000412; -.
DR PRINTS; PR00164; ABC2TRNSPORT.
DR PROSITE; PS00890; ABC2_MEMBRANE; 1.
KW Plasmid; Nodulation; Transport; Transmembrane; Inner membrane.
FT TRANSMEM 35 55
FT TRANSMEM 60 80
FT TRANSMEM 125 145
FT TRANSMEM 148 168
FT TRANSMEM 177 197
FT TRANSMEM 236 256
SQ SEQUENCE 262 AA; 28033 MW; F579E9EC3AE0A052 CRC64;

```

Query Match 1.2%; Score 7; DB 1; Length 262;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 TKAALAG 99
|||||
DB 123 TKAALAG 129

RESULT 40
ID GCTB_ACIFE STANDARD; PRT; 265 AA.
AC Q59112;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLUTACONATE COA-TRANSFERASE SUBUNIT B (EC 2.8.3.12) (GCT SMALL
DE SUBUNIT).
GN GCTB.
OS Acidaminococcus fermentans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Sporomusa subbranch;
OC Acidaminococcus.
OX NCBI_TaxID=905;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 43-65.
RC STRAIN=ATCC 25085;
RX MEDLINE=95045599; PubMed=7957258;
RA Mack M., Bendrat K., Zelder O., Eckel E., Linder D., Buckel W.;
RT "Location of the two genes encoding glutaconate coenzyme
RT A-transferase at the beginning of the hydroxyglutarate operon in
RT Acidaminococcus fermentans.";
RL Eur. J. Biochem. 226:41-51(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS).
RC STRAIN=ATCC 25085;
RX MEDLINE=97238937; PubMed=9083111;
RA Jacob U., Mack M., Clausen T., Huber R., Buckel W., Messerschmidt A.;
RT "Glutaconate Coa-transferase from Acidaminococcus fermentans: the
RT crystal structure reveals homology with other Coa-transferases.";
RL Structure 5:415-426(1997).

CC -1- FUNCTION: CATALYZES THE TRANSFER OF THE COA MOIETY FROM ACETYL COA
CC TO (R)-2-HYDROXYGLUTARATE AND RELATED COMPOUNDS LIKE GLUTACONATE.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + (E)-GLUTACONATE = ACETATE +
CC GLUTACONYL-1-COA.
CC -1- PATHWAY: GLUTAMATE FERMENTATION.
CC -1- SUBUNIT: HETEROOCTAMER OF FOUR A AND FOUR B SUBUNITS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLUTACONATE COA-TRANSFERASE SUBUNIT B
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; X81440; CAA57200.1;
CC PDB; 1POI; 18-MAR-98.
CC Transferase; 3D-structure.
CC INIT_MET 0
CC ACT_SITE 53 53
CC SEQUENCE 265 AA; 29035 MW; 1E7FF61B42162FB4 CRC64;
CC
CC Query Match 1.2%; Score 7; DB 1; Length 265;
CC Best Local Similarity 100.0%; Pred. No. 31;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 239 AVTIKQ 245
|||||

DB 13 AVTIKQ 19
RESULT 41
ID YN25_MYCTU STANDARD; PRT; 282 AA.
AC P71887;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 30.0 KDA PROTEIN RV2325C.
GN RV2325C OR MTCY3G12.09
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown S., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z79702; CAB02065.1;
CC TubercuList; RV2325C;
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 18 38 POTENTIAL.
CC TRANSMEM 40 60 POTENTIAL.
CC TRANSMEM 87 107 POTENTIAL.
CC TRANSMEM 119 139 POTENTIAL.
CC TRANSMEM 164 184 POTENTIAL.
CC TRANSMEM 260 280 POTENTIAL.
CC SEQUENCE 282 AA; 29955 MW; 689CDB75AA589D4E CRC64;

Query Match 1.2%; Score 7; DB 1; Length 282;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 AALAGGT 101
|||||
DB 98 AALAGGT 104

RESULT 42
ID YMX8_CAEEL STANDARD; PRT; 283 AA.
AC P34516;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 32.6 KDA PROTEIN K06H7.8 IN CHROMOSOME III.
GN K06H7.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Lathelle P., Lighning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; L15314; AAA28091.1; -;
 DR PIR; S44848; S44848.
 DR HSSP; Q06486; ICK1.
 DR WormPep; K06H7.8; CE00259.
 KW Hypothetical protein.
 SQ SEQUENCE 283 AA; 32597 MW; 84492C4D99984296 CRC64;

 Query Match 1.2%; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 368 KCVASGK 374
 Db 34 KCVASGK 40

 RESULT 43
 SM30_STRPU STANDARD; PRT; 290 AA.
 ID SM30_STRPU STANDARD; PRT; 290 AA.
 AC P28163;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 30 KDA SPICULE MATRIX PROTEIN PRECURSOR.
 GN SM30.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=92008838; PubMed=1717322;
 RA George N.C., Killian C.E., Wilt F.H.;
 RT "Characterization and expression of a gene encoding a 30.6-kDa
 RT Strongylocentrotus purpuratus spicule matrix protein.";
 RL Dev. Biol. 147:334-342(1991).
 CC -----
 CC FUNCTION: THE MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE. THE
 CC CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS.
 CC -----
 CC TISSUE SPECIFICITY: ACCUMULATES EXCLUSIVELY IN MINERALIZED
 CC TISSUES.

CC -----
 CC -!- DEVELOPMENTAL STAGE: IS INCREASED AT MIDDLE TO LATE MESENCHYME
 CC BLASTULA STAGE, LEVEL REMAINS HIGH THROUGH THE 3-DAY PLOTEUS
 CC STAGE.
 CC -----
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; M63840; AAA30070.1; -;
 DR InterPro; IPR001304; -;
 DR Pfam; PF00059; lectin_c; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; FALSE_NEG.
 DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 KW Matrix protein; Signal; Glycoprotein; Multigene family.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 290 30 KDA SPICULE MATRIX PROTEIN.
 FT DOMAIN 92 162 C-TYPE LECTIN.
 FT DOMAIN 271 281 ARG-RICH (BASIC).
 FT DOMAIN 215 219 POLY-ALA.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC...)(POTENTIAL).
 SQ SEQUENCE 290 AA; 31806 MW; E1E3E27E40796DF2 CRC64;

 Query Match 1.2%; Score 7; DB 1; Length 290;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 306 PVNPDPPT 312
 Db 39 PVNPDPPT 45

 RESULT 44
 SM3A_STRPU STANDARD; PRT; 290 AA.
 ID SM3A_STRPU STANDARD; PRT; 290 AA.
 AC Q26646;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 30 KDA SPICULE MATRIX PROTEIN-ALPHA PRECURSOR.
 GN SM30-ALPHA.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Sperm;
 RX MEDLINE=94327638; PubMed=80511158;
 RA Akasaka K., Frudakis T.N., Killian C.E., George N.C., Yamasu K.,
 RA Khaner O., Wilt F.H.;
 RT "Genomic organization of a gene encoding the spicule matrix protein
 RT SM30 in the sea urchin Strongylocentrotus purpuratus.";
 RL J. Biol. Chem. 269:20592-20598(1994).
 CC -----
 CC FUNCTION: THE MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE. THE
 CC CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS (BY SIMILARITY).
 CC -----
 CC TISSUE SPECIFICITY: ACCUMULATES EXCLUSIVELY IN MINERALIZED
 CC TISSUES (BY SIMILARITY).
 CC -----
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT THE TIME OF SPICULE FORMATION IN
 CC THE EMBRYO.
 CC -----
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U05962; AAB60620.1; -
DR EMBL; U05961; AAB60620.1; JOINED.
DR InterPro; IPR001304; -
DR Pfam; PF00059; lectin_c.1.
DR PROSITE; PS00615; C-TYPE_LLECTIN_1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE_LLECTIN_2; 1.
KW Matrix protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 290 30 KDA SPICULE MATRIX PROTEIN-ALPHA.
FT DOMAIN 92 162 C-TYPE LECTIN.
FT DOMAIN 24 27 POLY-GLY.
FT DOMAIN 215 220 POLY-ALA.
FT DOMAIN 271 281 ARG-RICH (BASIC).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 290 AA; 31686 MW; E89B5A5618E724D1 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 306 PVNPDPT 312
Db 39 PVNPDPT 45

RESULT 45
GLS1_BACSU STANDARD; PRT; 309 AA.
AC 007637;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE GLUTAMINASE YLAM (EC 3.5.1.2).
GN YLAM.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID:1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Purnelle B., Presecan E., Glaser P., Richou A., Danchin A.,
RA Goffeau A.;
RT "Bacillus subtilis chromosomal region downstream nprE.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + H(2)O = L-GLUTAMATE + NH(3).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINASE FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z97025; CAB09718.1; -
DR EMBL; Z99111; CAB13356.1; -
DR Subtilisin; BG13350; ylam.
KW Hydrolase.
SQ SEQUENCE 309 AA; 34012 MW; ELDFFFLA0723CIA5 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 309;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 151 ELEALVK 157

Db ||||| 14
8 ELEALVK 14
Search completed: July 30, 2001, 11:44:44
Job time: 13f sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2001, 11:42:06 ; Search time 21.49 Seconds
(without alignments)
3521.567 Million cell updates/sec

Title: US-09-367-496-8
Perfect score: 572
Sequence: 1 MSFOGKKSIPRITSRLIR.....RTAQKIMAPGGRSNIYSL 572

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Word size : 0
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL16:
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_unclassified.*
13: sp_vertebrate.*
14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	4.4	574	13 Q9DDZ6	Q9ddz6 gallus gall
2	14	2.4	519	11 Q9EQF5	Q9eqf5 mus musculus
3	14	2.4	589	5 Q9V3N7	Q9v3n7 drosophila
4	10	1.7	531	10 Q9FNP3	Q9fnp3 arabisidopsis
5	9	1.6	332	5 Q9VND9	Q9vnd9 drosophila
6	9	1.6	1030	5 Q9N445	Q9n445 caenorhabdi
7	8	1.4	319	2 Q9F3A7	Q9f3a7 streptomyce
8	8	1.4	376	10 Q9SBF7	Q9sbf7 arabisidopsis
9	8	1.4	487	2 Q52642	Q52642 pseudomonas
10	8	1.4	542	3 Q9P903	Q9p903 saccharomyc
11	8	1.4	563	11 Q9JMG8	Q9jmg8 rattus norv
12	8	1.4	564	4 Q9NRY9	Q9nry9 homo sapien
13	8	1.4	564	4 Q9NQC4	Q9nqc4 homo sapien
14	8	1.4	564	11 Q9JHU0	Q9jhu0 rattus norv
15	8	1.4	564	11 Q9EQF6	Q9eqf6 mus musculus
16	8	1.4	575	5 Q9NAP6	Q9nap6 caenorhabdi
17	8	1.4	581	2 Q9RJC2	Q9rjc2 streptomyce
18	8	1.4	648	10 Q9M269	Q9m269 arabisidopsis
19	8	1.4	657	13 Q9DER8	Q9der8 fugu rubrip

20	8	1.4	787	13 Q9DEC8	Q9dec8 oncorhynch
21	7	1.2	76	10 Q9LRF0	Q9lrf0 arabisidopsis
22	7	1.2	89	11 Q03317	Q03317 mus musculus
23	7	1.2	94	1 O28403	O28403 archaeoglob
24	7	1.2	112	2 Q9EZQ4	Q9ezq4 azoarcus sp
25	7	1.2	115	5 Q17259	Q17259 brachionus
26	7	1.2	115	14 Q9INE2	Q9ine2 human immun
27	7	1.2	119	5 Q9U4W3	Q9u4w3 plasmodium
28	7	1.2	123	2 Q9KB96	Q9kb96 bacillus ha
29	7	1.2	126	8 Q9TAJ8	Q9taj8 cafetaria r
30	7	1.2	127	2 Q9PMK0	Q9pmk0 campylobact
31	7	1.2	128	2 Q9LBT4	Q9lbt4 borrelia sp
32	7	1.2	129	2 Q9LBT3	Q9lbt3 borrelia sp
33	7	1.2	130	2 Q54544	Q54544 borrelia bu
34	7	1.2	130	2 Q54614	Q54614 borrelia bu
35	7	1.2	130	2 Q54501	Q54501 borrelia bu
36	7	1.2	130	2 Q54502	Q54502 borrelia bu
37	7	1.2	130	2 Q9LBT6	Q9lbt6 borrelia sp
38	7	1.2	135	2 Q9RVY5	Q9rvy5 deinococcus
39	7	1.2	143	2 Q9X5U7	Q9x5u7 borrelia bu
40	7	1.2	143	2 Q9RG11	Q9rg11 borrelia af
41	7	1.2	146	10 Q42474	Q42474 oryza sativ
42	7	1.2	147	14 Q98185	Q98185 molluscum c
43	7	1.2	151	2 Q31321	Q31321 borrelia bu
44	7	1.2	154	2 Q34675	Q34675 borrelia bu
45	7	1.2	155	2 Q31320	Q31320 borrelia bu

ALIGNMENTS

RESULT 1				
Q9DDZ6	PRELIMINARY;	PRT;	574 AA.	
ID Q9DDZ6				
AC Q9DDZ6				
DT 01-MAR-2001 (TREMREL. 16, Created)				
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)				
DE 01-MAR-2001 (TREMREL. 16, Last annotation update)				
DE COLLAPSED RESPONSE MEDIATOR PROTEIN 3.				
GN CRMP3.				
OS Gallus gallus (Chicken).				
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;				
OC Gallus.				
OX NCBI_TaxID=9031;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC TISSUE=RETINA;				
RX MEDLINE=20545548; PubMed=10956643;				
RA Fukada M., Watakabe I., Yuasa-Kawada J., Kawachi H., Kuroiwa A.,				
RA Matsuda Y., Noda M.;				
RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin				
RT Response Mediator Protein Family.";				
RL J. Biol. Chem. 275:37957-37965(2000).				
DR EMBL; AF249294; AAG37997.1; . 96E17AC0E96CA5F4 CRC64;				
SQ SEQUENCE 574 AA; 62070 MW; 96E17AC0E96CA5F4 CRC64;				

Query Match 4.4%; Score 25; DB 13; Length 574;
Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 GRFVPRKTFDFVYKRIKARNLAE 490				
Db 466 GRFVPRKTFDFVYKRIKARNLAE 490				
RESULT 2				
Q9EQF5	PRELIMINARY;	PRT;	519 AA.	
ID Q9EQF5				
AC Q9EQF5				
DT 01-MAR-2001 (TREMREL. 16, Created)				
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)				

```

Query Match      1.7%; Score 10; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      214 ELGITGPEGH 223
          |||||

```

Db 242 ELGTPGEGH 251

RESULT 5

Q9VND9 PRELIMINARY; PRT; 332 AA.
 AC Q9VND9; 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
 DE CRMP PROTEIN.
 GN CRMP OR BCDNA:HL02693 OR CG1411.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
 RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003602; AAF52003.2; -;
 DR FlyBase; FBgn0023023; CRMP.
 DR InterPro; IPR002195; -;
 DR Pfam; PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 332 AA; 36453 MW; 97BF4D2CFF3F601 CRC64;

Query Match 1.6%; Score 9; DB 5; Length 332;

Best Local Similarity 100.0%; Pred. No. 2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 PEVEAEAV 236

|||||||

Db 235 PEVEAEAV 243

RESULT 6

Q9N445 PRELIMINARY; PRT; 1030 AA.
 AC Q9N445; 01-OCT-2000 (TREMELrel. 15, Created)
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
 DE Y37ELLAL.C PROTEIN.
 GN Y37ELLAL.C.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC024757; AAF59446.1; -;
 DR InterPro; IPR000994; -;
 DR InterPro; IPR001714; -;
 DR Pfam; PF00557; Peptidase_M24; 1.
 DR PRINTS; PR00599; NAPEPTIDASE.
 SQ SEQUENCE 1030 AA; 115902 MW; 57058054CE8B60ED CRC64;

Query Match 1.6%; Score 9; DB 5; Length 1030;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 DGLIKQIGE 47

|||||||

Db 36 DGLIKQIGE 44

RESULT 7

Q9F3A7 PRELIMINARY; PRT; 319 AA.
 AC Q9F3A7; 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
 DE PUTATIVE REGULATORY PROTEIN.
 GN SC5F1.22C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,

```
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RM Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL450165; CAC16448.1; -.
SQ SEQUENCE 319 AA; 32573 MW; 3A35803084F8D8AC CRC64;

Query Match 1.4%; Score 8; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 559 MAPPGGRS 566
Db 171 MAPPGGRS 178
|||||

RESULT 8
Q9SBF7 PRELIMINARY; PRT; 376 AA.
AC Q9SBF7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR.
GN MYB68.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=99056848; PubMed=9839469;
RA Kranz H.D., Denekamp M., Greco R., Jin H.-L., Leyva A., Weissner R.,
RA Petroni K., Urzaingui A., Bevan M., Martin C., Smeekens S.,
RA Tonelli C., Paz-Ares J., Weisshaar B.;
RT "Towards functional characterisation of the members of the R2R3-MYB
RT gene family from Arabidopsis thaliana.";
RL Plant J. 16:263-276(1998).
DR EMBL; AF062901; AAC83623.1; -.
DR HSSP; P01103; 1POM.
DR InterPro; IPR001005; -.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR PROSITE; PS00037; MYB.1; UNKNOWN_1.
DR PROSITE; PS00334; MYB.2; UNKNOWN_1.
DR PROSITE; PS00090; MYB.3; 2.
DR SMART; SM00395; SANT; 1.
SQ SEQUENCE 376 AA; 42084 MW; 342FF295D5D5DED22 CRC64;

Query Match 1.4%; Score 8; DB 10; Length 376;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 AALAGGTT 102
Db 337 AALAGGTT 344
|||||

RESULT 9
O52642 PRELIMINARY; PRT; 487 AA.
ID O52642;
AC O52642;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE NITROBENZALDEHYDE DEHYDROGENASE NTNC.
GN NTNC.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=TW3;
RC MEDLINE=982151569; PubMed=9555884;
RX James K.D., Williams P.A.;
RA "ntn genes determining the early steps in the divergent catabolism of
RT 4-nitrotoluene and toluene in Pseudomonas sp. strain TW3.";
RL J. Bacteriol. 180:2043-2049(1998).
DR EMBL; AF043544; AAC38357.1; -.
DR HSSP; P56533; 1BPW.
DR InterPro; IPR002086; -.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
SQ SEQUENCE 487 AA; 51806 MW; 6E2FDAF4BC4B8497 CRC64;

Query Match 1.4%; Score 8; DB 2; Length 487;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 VLPGGVDV 72
Db 208 VLPGGVDV 215
|||||

RESULT 10
Q9P903 PRELIMINARY; PRT; 542 AA.
ID Q9P903;
AC Q9P903;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DIHYDROPYRIMIDINASE (EC 3.5.2.2).
GN PYD2.
OS Saccharomyces kluyveri (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4934;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-12651;
RX MEDLINE=20123992; PubMed=10656811;
RA Gojkovic Z., Jahneke K., Schnackerz K.D., Piskur J.;
RT "PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates
RT in a novel fungal catabolic pathway.";
RL J. Mol. Biol. 295:1073-1087(2000).
DR EMBL; AF158967; AAF69237.1; -.
DR InterPro; IPR002195; -.
DR Pfam; PF00744; Dihydroorotase; 1.
KW Hydrolase.
SQ SEQUENCE 542 AA; 60240 MW; E8758452CBE86E98 CRC64;

Query Match 1.4%; Score 8; DB 3; Length 542;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 SDADLVIV 412
Db 446 SDADLVIV 453
|||||

RESULT 11
Q9JMG8 PRELIMINARY; PRT; 563 AA.
ID Q9JMG8;
AC Q9JMG8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DIHYDROPYRIMIDINASE-RELATED PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yanagi S., Inatome R., Yamamura H.;
 RT "Dihydropyrimidinase-related protein.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB029432; BAA89475.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 563 AA; 61395 MW; B36767BF1440004D CRC64;

Query Match 1.4%; Score 8; DB 11; Length 563;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 207 LGITGPEG 214

RESULT 12

O9NRY9 PRELIMINARY; PRT; 564 AA.
 AC O9NRY9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE COLLAPSED RESPONSE MEDIATOR PROTEIN-5.
 GN CRMP5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yu Z., Kyzer T.J., Griesmann G.E., Lennon V.A.;
 RT "Collapsin response mediator protein-5 (CRMP5), a novel autoantigen
 associated with paraneoplastic neurological disorders.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF157634; AAF80348.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61459 MW; 949947A983735340 CRC64;

Query Match 1.4%; Score 8; DB 4; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 13

O9NQC4 PRELIMINARY; PRT; 564 AA.
 AC O9NQC4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 61.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SPINAL CORD;
 RA Horiuchi M., Betz H.;
 RT "Human homologue of Ulip6.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ251275; CAB95124.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 564 AA; 61394 MW; F370DD35B0B49935 CRC64;

Query Match 1.4%; Score 8; DB 4; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 14

O9JHU0 PRELIMINARY; PRT; 564 AA.
 AC O9JHU0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE ULIP-LIKE PROTEIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Horiuchi M., Betz H.;
 RT "Ulip and dihydropyrimidinase like protein.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ131436; CAB95193.1; -
 DR InterPro: IPR002195; -
 DR Pfam: PF00744; Dihydroorotase; 1.
 SQ SEQUENCE 564 AA; 61540 MW; E078B9002F54975E CRC64;

Query Match 1.4%; Score 8; DB 11; Length 564;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 LGITGPEG 222
 |||||
 DB 208 LGITGPEG 215

RESULT 15

O9EQF6 PRELIMINARY; PRT; 564 AA.
 AC O9EQF6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE COLLAPSED RESPONSE MEDIATOR PROTEIN 5.
 GN CRMP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SWISS WEBSTER/NIH; TISSUE-WHOLE EMBRYO;
 EX MEDLINE-20545548; PubMed-10956643;
 RA Fukada M., Wataabe I., Yuasa Kawada J., Kawachi H., Kuroiwa A.,
 RA Matsuda Y., Noda M.;
 RT "Molecular Characterization of CRMP5, a Novel Member of the Collapsin
 Response Mediator Protein Family.";
 RL J. Biol. Chem. 275:37957-37965(2000).
 DR EMBL: AF249295; AAG37998.1; -
 SQ SEQUENCE 564 AA; 61516 MW; CA93790FC8F9CD98 CRC64;

```

Query Match          1.4%; Score 8; DB 11; Length 564;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 LGITGPEG 222
Db 208 LGITGPEG 215
|||||||

RESULT 16
Q9NAP6 PRELIMINARY; PRT; 575 AA.
AC Q9NAP6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE K09E4.4 PROTEIN.
GN K09E4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; 283234; CAB70170.1; -.
SQ SEQUENCE 575 AA; 66849 MW; FE7216DB39F4D93C CRC64;

Query Match          1.4%; Score 8; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 RLLELGIT 218
Db 98 RLLELGIT 105
|||||||

RESULT 17
Q9RJG2 PRELIMINARY; PRT; 581 AA.
AC Q9RJG2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 63.8 KDA PROTEIN.
GN SCF91.02C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL132973; CAB61159.1; -.
DR InterPro; IPR001440; -.
DR Pfam; PF00515; TPR; 3.
KW Hypothetical protein.
SQ SEQUENCE 581 AA; 63847 MW; 65F1A06AE3D53C CRC64;

Query Match          1.4%; Score 8; DB 2; Length 581;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 EAEAVYRA 239
Db 494 EAEAVYRA 501
|||||||

RESULT 18
Q9M269 PRELIMINARY; PRT; 648 AA.
AC Q9M269;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE-LIKE PROTEIN.
GN F21F14.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Altiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quefier F., Salanoubat M.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL138642; CAB71903.1; -.
DR InterPro; IPR000719; -.
DR InterPro; IPR001245; -.
DR InterPro; IPR002290; -.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR SMART; SM00220; S_TK; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase;
KW Tyrosine-protein kinase.
SQ SEQUENCE 648 AA; 72131 MW; 20808421BF6A3C55 CRC64;

Query Match          1.4%; Score 8; DB 10; Length 648;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 VNPDPPTA 314
Db 515 VNPDPPTA 522
|||||||

RESULT 19

```


Tue Jul 31 13:08:34 2001

us-09-367-496-8.oli.rspt

```

Q9DER8
ID Q9DER8 PRELIMINARY; PRT; 657 AA.
AC Q9DER8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSPORT ASSOCIATED PROTEIN.
GN ABCB3B.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20491935; PubMed=11035083;
RA Clark M.S., Pontarotti P., Gilles A., Kelly A., Elgar G.;
RT "Identification and characterization of a beta proteasome subunit
cluster in the Japanese pufferfish (Fugu rubripes).";
RL J. Immunol. 165:4446-4452(2000).
DR EMBL; AJ271723; CAC13121.1; -.
SQ SEQUENCE 657 AA; 72750 MW; B657D0CAAF61945E CRC64;

Query Match 1.4%; Score 8; DB 13; Length 657;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 VVISQGRV 452
DB 612 VVISQGRV 619

RESULT 20
Q9DEC8 PRELIMINARY; PRT; 787 AA.
ID Q9DEC8;
AC Q9DEC8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE COMPLEMENT FACTOR B/C2-B.
GN BFC2-B.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RA Nakao M., Fujiki K., Yano T., Bayne C.J.;
RT "Molecular cloning of B/C2-B, an isotype of complement factor B/C2,
from teleost.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB044939; BAB19788.1; -.
SQ SEQUENCE 787 AA; 87578 MW; 9111C72F80816B0C CRC64;

Query Match 1.4%; Score 8; DB 13; Length 787;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 SGFSLSGS 544
DB 136 SGFSLSGS 143

RESULT 21
Q9LFF0 PRELIMINARY; PRT; 76 AA.
ID Q9LFF0;
AC Q9LFF0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)

Q9DER8
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 8.2 KDA PROTEIN.
GN FSE19.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391147; CAC01831.1; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8169 MW; F60F2CA8BBEF137A CRC64;

Query Match 1.2%; Score 7; DB 10; Length 76;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 EYEAFAV 236
DB 62 EYEAFAV 68

RESULT 22
Q03317 PRELIMINARY; PRT; 89 AA.
ID Q03317;
AC Q03317;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 9.3 KDA PROTEIN (ORF1).
GN TCEAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83153686; PubMed=3346229;
RA Kanai A., Kuzuhara T., Sekimizu K., Natori S.;
RT "Molecular cloning and characterization of cDNA for eukaryotic
transcription factor S-II.";
RL J. Biochem. 109:674-677(1991).
DR EMBL; D00925; BAA00767.1; -.
DR MGD; MGI:1196624; Tceal.
KW Hypothetical protein.
SQ SEQUENCE 89 AA; 9348 MW; 3EF3976B06857986 CRC64;

Query Match 1.2%; Score 7; DB 11; Length 89;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 OGRVALE 455
DB 66 OGRVALE 72

RESULT 23
Q28403 PRELIMINARY; PRT; 94 AA.
ID Q28403;
AC Q28403;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

```

DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF1876.

OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervatage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE000973; AAB89376.1; -.

DR TIGR; AF1876; -.
KW Hypothetical protein.
SQ SEQUENCE 94 AA; 10657 MW; FDD6FB00B313918E CRC64;

Query Match 1.2%; Score 7; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LVKEGV 161
Db 85 LVKEGV 91
|||||

RESULT 24
Q9EZQ4
ID Q9EZQ4 PRELIMINARY; PRT; 112 AA.
AC Q9EZQ4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE P2-LIKE SIGNAL TRANSMITTER PROTEIN GLNY.
GN GLNY.
OS Azoarcus sp. BH72.
OC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
OC Azoarcus.
OX NCBI_TaxID=62928;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BH72;
RX MEDLINE=20521930; PubMed=11069654;
RA Martin D.E., Hurek T., Reinhold-Hurek B.;

RT "Occurrence of three PII-like signal transmitter proteins in the
diatrophic proteobacterium Azoarcus sp. BH72.";
RL Mol. Microbiol. 38:276-288(2000).
DR EMBL; AF281015; AAG40186.1; -.
SQ SEQUENCE 112 AA; 12150 MW; C18A32878387C16A CRC64;

Query Match 1.2%; Score 7; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 DGKMFVT 462
Db 88 DGKMFVT 94
|||||

RESULT 25
Q17259
ID Q17259 PRELIMINARY; PRT; 115 AA.
AC Q17259;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CHAPERONIN (FRAGMENT).
GN HSP60.
OS Brachionus plicatilis.
OC Eukaryota; Metazoa; Rotifera; Monogononta; Plolmida; Brachionidae;
OC Brachionus.
OX NCBI_TaxID=10195;
RN [1]
RP SEQUENCE FROM N.A.
RL Cochran B.J., Mattley Y.M., Snell T.W.;

RL Environ. Toxicol. Chem. 13:1221-1229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Cochran B.J.;

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20804; AAA64304.1; -.
DR InterPro; IPR002423; -.
DR Pfam; PF00118; cph60_TCP1; 1.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 12697 MW; 5A24CD3F43B87CDE CRC64;

Query Match 1.2%; Score 7; DB 5; Length 115;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EEVEAEA 235
Db 84 EEVEAEA 90
|||||

RESULT 26
Q9INE2
ID Q9INE2 PRELIMINARY; PRT; 115 AA.
AC Q9INE2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DRCM1.
RX MEDLINE=99260287; PubMed=10331444;
RA Mokill J.L., Wade C.M., Burns S.M., Cutting W.A., Bopopi J.M.,
RA Green S.D., Peutherer J.F., Simmonds P.;

RT "Genetic heterogeneity of HIV type 1 subtypes in Kinshasa, rural
Democratic Republic of Congo.";
RL AIDS Res. Hum. Retroviruses 15:655-664(1999).
CC -1- MISCELLANEOUS: GAG GENE PROTEIN P17 FOUND N-TERMINAL TO GAG P24
(CORE NUCLEOCAPSID PROTEIN) IN THE GAG POLYPROTEIN (BY
SIMILARITY).
CC -1- SIMILARITY: TO RETROVIRAL MATRIX PROTEIN (GAG GENE PROTEIN P17).
CC EMBL; AF144807; AAF69033.1; -.
DR InterPro; IPR000071; -.
DR Pfam; PF00540; gag_p17; 1.
DR PRINTS; PR00234; HIVMATRIX.
KW AIDS: Core protein; Polyprotein.
FT NON_TER 1
FT NON_TER 115
SQ SEQUENCE 115 AA; 12712 MW; 83A6FAF50A9CC648 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 DGKMFVT 462
Db 88 DGKMFVT 94
|||||

RESULT 24
Q9EZQ4
ID Q9EZQ4 PRELIMINARY; PRT; 112 AA.
AC Q9EZQ4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE P2-LIKE SIGNAL TRANSMITTER PROTEIN GLNY.
GN GLNY.
OS Azoarcus sp. BH72.
OC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
OC Azoarcus.
OX NCBI_TaxID=62928;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BH72;
RX MEDLINE=20521930; PubMed=11069654;
RA Martin D.E., Hurek T., Reinhold-Hurek B.;

RT "Occurrence of three PII-like signal transmitter proteins in the
diatrophic proteobacterium Azoarcus sp. BH72.";
RL Mol. Microbiol. 38:276-288(2000).
DR EMBL; AF281015; AAG40186.1; -.
SQ SEQUENCE 112 AA; 12150 MW; C18A32878387C16A CRC64;

QY 208 EQRLLLE 214
 Db 85 EQRLLLE 91
 |||||
 Db 37 SLGSGQA 43
 |||||
 RESULT 31
 Q9LBT4
 ID Q9LBT4 PRELIMINARY; PRT; 128 AA.
 AC Q9LBT4;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia sp. TMI.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=113255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TMI;
 RA Masuzawa T., Kudeken M., Ochi A.;
 FT "Borrelia TMI flagellin gene."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB037129; BAA90306.1; -
 DR InterPro; IPR001444; -
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 128 128
 FT NON_TER 128 128
 SQ SEQUENCE 128 AA; 13114 MW; E31B76B94478BC72 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 540 SLGSGQA 546
 |||||
 Db 37 SLGSGQA 43

RESULT 32
 Q9LBT3
 ID Q9LBT3 PRELIMINARY; PRT; 129 AA.
 AC Q9LBT3;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia sp. TAL.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=113256;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TAL;
 RA Masuzawa T., Kudeken M., Ochi A.;
 FT "Borrelia TAL flagellin gene."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB037130; BAA90307.1; -
 DR InterPro; IPR001444; -
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 129 129
 FT NON_TER 129 129
 SQ SEQUENCE 129 AA; 13212 MW; 26B03A44B94478BC CRC64;

Query Match 1.2%; Score 7; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 540 SLGSGQA 546
 |||||

Db 37 SLGSGQA 43
 |||||
 RESULT 33
 O54544
 ID O54544 PRELIMINARY; PRT; 130 AA.
 AC O54544;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,
 RA Black W.C. IV;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U96235; AAC09048.1; -
 DR EMBL; U96235; AAC09047.1; -
 DR InterPro; IPR001444; -
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 130 130
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13723 MW; 711326BFB9C05E7 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 540 SLGSGQA 546
 |||||
 Db 63 SLGSGQA 69

RESULT 34
 O54614
 ID O54614 PRELIMINARY; PRT; 130 AA.
 AC O54614;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE FLAGELLIN PROTEIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Norris D.E., Johnson B.J.B., Piesman J., Maupin G.O., Clark J.L.,
 RA Black W.C. IV;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U96238; AAC09050.1; -
 DR EMBL; U96237; AAC09049.1; -
 DR InterPro; IPR001444; -
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 FT NON_TER 130 130
 FT NON_TER 130 130
 SQ SEQUENCE 130 AA; 13779 MW; EF0BAFBA395AB2 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 540 SLGSGQA 546
 |||||

	Query Match	1.2%	Score 7;	DB 2;	Length 130;	
	Best Local Similarity	100.0%;	Pred. No. 94;			
	Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
	Query Match	1.2%;	Score 7;	DB 2;	Length 135;	
	Best Local Similarity	100.0%;	Pred. No. 98;			
	Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
SQ	SEQUENCE	135 AA;	14523 MW;	Cf33540LCPAR16002	G86347	

QY 65 VLPGGVD 71
Db 101 VLPGGVD 107

RESULT 39
Q9X5U7
ID Q9X5U7 PRELIMINARY; PRT; 143 AA.
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLAGELLIN PROTEIN (FRAGMENT).
OS Borrelia burgdorferi group.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=64895;
RN [1]
RP SEQUENCE FROM N.A.
RA Leutenegger C.M., Pusterla N., Mislin C., Lutz H.;
RT "Molecular evidence of Ticks infected with Borrelia burgdorferi sensu
latu in Switzerland."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF127531; AAD31190.1; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
DR PROSITE; PS00588; FLAGELLIN_BB_ROD; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 143
SQ SEQUENCE 143 AA; 14913 MW; 733C3EB0171C4C8B CRC64;

Query Match 1.2%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLGSGQA 546
Db 44 SLGSGQA 50

RESULT 40
Q9RG11
ID Q9RG11 PRELIMINARY; PRT; 143 AA.
AC Q9RG11
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLAGELLIN PROTEIN (FRAGMENT).
OS Borrelia afzelii
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29518;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99419177; PubMed=10488215;
RA Leutenegger C.M., Pusterla N., Mislin C.N., Weber R., Lutz H.;
RT "Molecular evidence of coinfection of ticks with Borrelia burgdorferi
sensu lato and the human granulocytic ehrlichiosis agent in
Switzerland."
RL J. Clin. Microbiol. 37:3390-3391(1999).
DR EMBL; AF127532; AAD31191.1; -
DR InterPro; IPR001444; -
DR Pfam; PF00460; flg_bb_rod; 1.
FT NON_TER 1
FT NON_TER 143
SQ SEQUENCE 143 AA; 14872 MW; 98A5BB44B2972B0C CRC64;

Query Match 1.2%; Score 7; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 540 SLGSGQA 546
Db 44 SLGSGQA 50

Db 44 SLGSGQA 50

RESULT 41
Q42474
ID Q42474 PRELIMINARY; PRT; 146 AA.
AC Q42474
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ROOT-SPECIFIC PROTEIN (PUTATIVE LIPID TRANSFER PROTEIN).
GN RCC2 OR RCC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-GULFMONT; TISSUE-ROOT;
RX MEDLINE=95195153; PubMed=7888615;
RA Xu Y., Buchholz W.G., DeRose R.T., Hall T.C.;
RT "Characterization of a rice gene family encoding root-specific
proteins."
RL Plant Mol. Biol. 27:237-248(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. NIPPONBARE;
RA Hsiao C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Buiac J., Zisman V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feildilyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBa0015J15 genomic sequence."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; L27210; AAA79836.1; -
DR EMBL; L27209; AAA65512.1; -
DR EMBL; AC026758; AAG13492.1; -
DR HSP; P24337; 1HYP.
DR Mendel; 9716; Oryza; 1531; 9716.
DR InterPro; IPR000528; -
DR Pfam; PF00279; LTP; 1.
SQ SEQUENCE 146 AA; 14466 MW; E51E2E89F193457F CRC64;

Query Match 1.2%; Score 7; DB 10; Length 146;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 441 GAPAVVI 447
Db 32 GAPAVVI 38

RESULT 42
Q98185
ID Q98185 PRELIMINARY; PRT; 147 AA.
AC Q98185; 012881; 012600;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MC014R.
GN MC014R OR B-M,N,L.4.
OS Mollusca; Contagiosum virus subtype 1 (MCV1).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Molluscipoxvirus.
OX NCBI_TaxID=10280;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96325459; PubMed=8670425;
RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RT "Genome sequence of a human tumorigenic poxvirus: prediction of
specific host response-evasion genes.";

* Tue Jul 31 13:08:34 2001

RL Science 273:813-816(1996).
 RP SEQUENCE FROM N.A.
 RA Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
 RA Moss B.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97093414; PubMed=8938976;
 RA Martin-Gallardo A., Moratilla M., Funes J.M., Agronayor M., Nunez A.,
 RA Varas A.J., Collado M., Valencia A., Lopez-Esteban J.L.,
 RA Esteban M.;
 RT "Sequence analysis of a Molluscum contagiosum virus DNA region which
 RT includes the gene encoding protein kinase 2 and other genes with
 RT unique organization.";
 RL Virus Genes 13:19-29(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Moratilla M., Agronayor M., Nunez A., Funes J.M., Varas A.J.,
 RA Lopez-Esteban J.L., Esteban M., Martin-Gallardo A.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Moratilla M., Agronayor M., Nunez A., Funes J.M., Varas A.J.,
 RA Lopez-Esteban J.L., Collado M., Pavon M., Martin-Gallardo A.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U60315; AAC55142.1; -;
 DR EMBL; U60315; AAC55142.1; -;
 DR EMBL; U60315; AAC55142.1; -;
 DR EMBL; U60315; AAC55142.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 147 AA; 16225 MW; AF6750D93EBA2F1B CRC64;

Query Match 1.2%; Score 7; DB 14; Length 147;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 APPGGRS 566
 DB 120 APPGGRS 126

RESULT 43
 ID 031321 PRELIMINARY; PRT; 151 AA.
 AC 031321;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-UK-SCO;
 RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y15100; CAA75375.1; -;
 DR InterPro; IPR001029; -;
 DR InterPro; IPR001444; -;
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 DR NON_TER 1
 FT NON_TER 151
 FT SEQUENCE 151 AA; 15697 MW; 4B28C6E0A0ACC9E CRC64;

Query Match 1.2%; Score 7; DB 2; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSCA 546
 DB 6 SLSGSCA 12
 RESULT 44
 ID 034675 PRELIMINARY; PRT; 154 AA.
 AC 034675;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IP3, AND CHIN23;
 RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y15095; CAA75370.1; -;
 DR EMBL; Y15095; CAA75370.1; -;
 DR InterPro; IPR001029; -;
 DR InterPro; IPR001444; -;
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 DR NON_TER 1
 FT NON_TER 154
 FT SEQUENCE 154 AA; 16078 MW; 69BA0FC908EC8119 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 SLSGSCA 546
 DB 7 SLSGSCA 13

RESULT 45
 ID 031320 PRELIMINARY; PRT; 155 AA.
 AC 031320;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE FLAGELLIN (FRAGMENT).
 GN FLA.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-UK-OX;
 RA Curtin S.M., Maggs A.D.F., Carter P.E., Pennington T.H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y15099; CAA75374.1; -;
 DR InterPro; IPR001029; -;
 DR InterPro; IPR001444; -;
 DR Pfam; PF00460; flg_bb_rod; 1.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 DR NON_TER 1
 FT NON_TER 155
 FT SEQUENCE 155 AA; 16058 MW; 6BAB1FB4CA30BEE7 CRC64;

Query Match 1.2%; Score 7; DB 2; Length 155;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Tue Jul 31 13:08:34 2001

us-09-367-496-8.oli.rspt

Page 14

Matches	7;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	540	SLGSQA	546						
Db	7	SLGSQA	13						

Search completed: July 30, 2001, 11:44:20
Job time: 134 sec